Contract No.:   500-94-0062
MPR Reference No.: 8246

Preliminary Report: The Impact of Prospective Payment on Medicare Home Health Use-- Promising Results for a Future Program

July 22, 1998

Valerie Cheh
Christopher Trenholm
Randall Brown
Barbara Phillips

CONTENTS

CONTENTS *(continued)*

TABLES

FIGURES

EXECUTIVE SUMMARY

As part of its ongoing effort to study methods of providing more cost-effective care, the Health Care Financing Administration (HCFA) has recently implemented the Per-Episode Home Health Prospective Payment Demonstration. Under the demonstration, participating home health agencies are paid a fixed, lump-sum payment for the first 120 days of each episode of care provided to Medicare beneficiaries and a predetermined rate for each visit thereafter. This method of compensation differs substantially from the current method of Medicare reimbursement for home health services, under which agencies are reimbursed for actual costs incurred, up to a specific limit. By allowing agencies to retain most of any surplus payments over cost, prospective payment gives agencies a financial incentive to provide home health care in a more cost-efficient manner than under traditional cost-based reimbursement.

Ninety-one agencies in five states entered the three-year demonstration at the start of their 1996 fiscal year. Prior to the start of the demonstration, the participating agencies were randomly assigned to either the treatment group (which is paid under the demonstration's prospective payment method) or a control group (which continues to be paid under Medicare's normal method of cost-based reimbursement). The payments treatment group agencies receive for the first 120 days of a patient episode are based on each agency's own costs in the fiscal year immediately preceding its entry into the demonstration, adjusted for changes in its case mix. While each agency is "at risk" during the first 120 days after admission for all home health visits the patient needs, HCFA reimburses treatment agencies for up to 99 percent of fiscal-year losses up to the Section 223 payment limits.[1] Profits in excess of specified limits must be shared with HCFA.

RESEARCH QUESTIONS AND METHODOLOGY

In this report, we examine the available data from (roughly) the first year of the demonstration to test hypotheses about the possible effects of prospective payment on patients' use of Medicare-covered services. Given the limited data currently available, this preliminary report focuses only on home health use taking place during the "at-risk" period (first 120 days) of a home health episode. As more data become available, future reports will provide a more complete investigation of demonstration impacts on home health use and other outcomes. Here, we test hypotheses concerning the impacts of the demonstration on (1) the number of visits provided, both in total and by type; (2) the per-episode mix of services provided; and (3) the length of an episode of care. In addition, we also test whether these outcomes differed between subgroups of agencies defined by their for-profit status, size, and other key characteristics.

The analysis is based on approximately 51,000 home health episodes taking place in 85 of the demonstration agencies (6 of the 91 agencies were excluded because they dropped out of the

---

[1]The Section 223 payment limits are cost-per-visit payment limits that apply to all agencies in the Medicare program.

demonstration or had inadequate data). All admissions occurring between an agency's start date and August 1996 are included. Medicare claims files provided the data on the outcome variables describing the use of services during the first 120 days of home health episodes. Data collected at admission for case-mix adjustment and from prior Medicare claims provided measures of preadmission characteristics of patients admitted to agencies participating in the demonstration. Data on agency characteristics were obtained from the agency cost reports and the demonstration implementation contractor.

Ordinary least squares models and logistic models were used to estimate program effects, controlling for preexisting differences between treatment and control agencies in patient and agency characteristics. This approach proved crucial to obtaining valid impact estimates since, despite the randomization of participating agencies, there were several significant differences between treatment and control agencies aside from the method of payment. Observations are weighted so that each agency is represented equally in the analysis. Standard errors of impact estimates were calculated using special software designed to account for the effects of sample clustering and weighting, so as to avoid overstating the precision of the estimates. Analyses of the robustness of our regression estimates showed that they were not sensitive to the weighting scheme, statistical methods, or model specifications used.

FINDINGS

Visits per Episode Fell 17 Percent

We find strong evidence that prospective payment reduced the total number of visits per episode, with large and statistically significant effects on skilled nursing, home health aide, and medical social worker visits. In addition, we find relatively large but statistically insignificant declines in the number of visits by physical and occupational therapists. Overall, we estimate that prospective payment led treatment agencies to provide an average of 7.8 fewer visits than control agencies during the first 120 days of each episode, a decline of 17 percent relative to the mean for control group agencies (see Figure 1). This includes an estimated per-episode reduction of about 4 skilled nursing visits (down 18 percent), 2.7 home health aide visits (down 18 percent), 0.6 physical therapy visits (down 10 percent), and 0.3 medical social worker visits (down 37 percent). The effects on skilled nursing and aide visits dominate the effects on overall visits because they account for 83 percent of all visits during the first 120 days of an episode.

FIGURE 1

AVERAGE NUMBER OF VISITS PER EPISODE
BY TREATMENT AND CONTROL AGENCIES
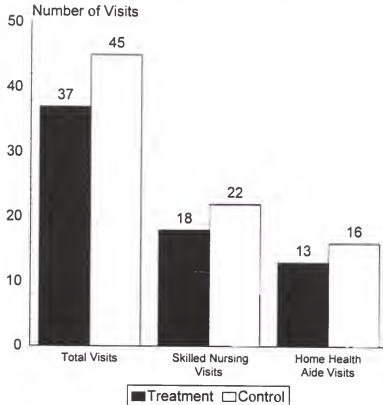
Source: Medicare Claims Data.

Note: Treatment group mean is adjusted to account for preexisting differences between treatment and control agencies.

The Percent Receiving Occupational Therapy Declined

The proportion of patients receiving occupational therapy declined by about one-third from the control group mean of 12.6 percent, falling by 4.4 percentage points. However, the proportion receiving other services was relatively unaffected. The reduction in occupational therapy may be due to fewer patients receiving assessment visits from occupational therapists. The program effect on the receipt of any skilled nursing visits, a drop of 2.6 percentage points, was also statistically significant, but small. This reduction may be due to agencies having some cases opened by physical therapists, when therapy is the primary service the patient needs. For the other services--aide care (received by 46 percent of patients), physical therapy (42 percent), speech therapy (3 percent), and medical social worker visits (29 percent)--there were no discernible effects.

Prospective Payment Had No Major Effect on the Mix of Visits Provided

Prospective payment appeared to have no effect on the proportion of visits per episode accounted for by any particular specialty. Estimated effects on the proportions of episode visits that were for skilled nursing, therapies (physical, occupation, and speech combined), aide care, and medical social services were each small and statistically insignificant, suggesting that the impacts of the demonstration fell roughly proportionally across the major visit types.

Episodes Were Shortened by 14 Percent

Consistent with the impacts on the number of visits provided, prospective payment reduced the length of episodes within the at-risk period by about 10 days, or 14 percent of the mean for control group agencies (70 days). Moreover, prospective payment reduced the probability that an episode exceeds 120 days by about 30 percent (10 percentage points) relative to the control group mean of 35 percent. This effect, which accounts for the majority of the total reduction in average episode length, suggests that the proportion of long home health episodes may be significantly reduced. The finding also suggests that, when service use during the period after 120 days is examined, the total reduction in visits due to prospective payment may be substantially larger. However, it will also be necessary to assess whether treatment group patients have more readmissions.

Program Impacts Varied Little with Agency Characteristics

We tested whether impacts on the following outcomes varied with agency characteristics: (1) number of visits per episode, (2) the proportion of visits that were skilled nursing visits, (3) the proportion of visits that were aide visits, and (4) episode length. Agency characteristics used to define subgroups included whether agencies were for-profit, their size, whether they were freestanding or hospital-based, and their predemonstration practice pattern (that is, whether the agency provided more or less than the average number of visits per episode during the predemonstration year, for the case mix served).

We find only one important difference in impacts across the agency subgroups. The impact on total visits was significantly greater for agencies with high-use prior practice patterns than for those with low-use patterns. High-use agencies cut visits per episode by nearly 11 visits (about 20 percent of the control group mean), compared to a drop of only about 5 visits (14 percent) for agencies with low-use practice patterns.

Agency size and auspices (freestanding or hospital-based) have no apparent influence on the estimated effects for any of the outcomes examined, but effects on the mix of visits may differ by agencies' for-profit status. For-profit agencies significantly reduced the proportion of skilled nursing visits per episode, but nonprofit agencies did not. Effects on other outcomes were similar for for-profit and nonprofit agencies. Given that for-profit agencies tended to have higher-use practice patterns in the predemonstration period, program effects on visits per episode may be greater on average for for-profit agencies than for nonprofit ones.

CONCLUSIONS

Prospective Payment May Yield Substantial Cost Savings and Increased Flexibility for Agencies

Our findings suggest that prospective payment for home health agencies could lead to potentially sizable savings in costs, depending on how the payment rates are set. The 17 percent reduction in visits, however, does not necessarily imply similar size reductions in costs, because agencies' cost per visit may increase if their total volume of visits declines. Nonetheless, if savings of roughly this magnitude are achievable once agencies adjust to a smaller scale or consolidate, the potential exists both for Medicare to save and for agencies to prosper. Savings to Medicare could be achieved by setting payment rates at a proportion (for example, 90 percent) of the average expected cost per episode under cost-based reimbursement, as the Medicare risk program does for payments to health maintenance organizations.

The finding that nearly all types of visits appear to be reduced suggests that it could be beneficial for HCFA to design its future prospective payment system in a way that gives agencies the flexibility to choose which services to decrease. In the Balanced Budget Act of 1997, Congress restricted coverage for skilled nursing for patients needing blood drawn. This effort to reduce home health utilization focused on decreasing the use of skilled nursing and home health aide services. Decreasing the use of these services is logical because these services are used most frequently and, thus, are thought to provide the greatest opportunities for cost savings. However, this analysis shows that, given the right incentives, agencies may also make substantial reductions in therapy and medical social worker visits, which account for about 17 percent of all visits in the first 120 days of an episode. Therefore, a prospective payment system that encompasses all services rendered could be more beneficial to both the government and home health agencies.

The observed reduction in episode length is also potentially important for designing the future system. The 120-day period examined here is somewhat shorter than the 150-day episode length discussed recently before Congress as being necessary to encompass 80 percent of Medicare home health patients. On the basis of our estimates of the reduction in long episodes, it appears that a 120-day at-risk period would still capture about 80 percent of all Medicare home health patients. Thus, HCFA may wish to consider using a shorter at-risk period than the experience under cost-based reimbursement would suggest.

These Results Are Promising but Preliminary

While the results are promising, they must be viewed as preliminary, for several reasons. First, we must conduct further analyses to ensure that the results are not due to underreporting of visits by agencies paid prospectively. While our preliminary analysis suggests that this is not likely to be true, further investigation is warranted. Second, the data used are from only the first 8 to 15 months of the demonstration. Impacts may change as agencies become more adept at finding ways to provide home health care efficiently. Third, these findings provide no information on the *consequences* of the observed reduction in the number of visits per episode. Future analyses will examine the effects of the demonstration on patient outcome measures that reflect the quality of care provided. Fourth, the potential cost savings will be affected by demonstration effects on the cost per visit and by

effects on service use in the period after the first 120 days of an episode. Effects on potential costs to HCFA also depend on whether beneficiaries' use of other Medicare-covered services is affected. In addition, we will want to explore whether the reductions in visits per episode seem to be greater for some types of beneficiaries than for others.

Even when the final estimates are obtained, generalizing from the demonstration results to the expected effects of a national program must be done with caution. A national program would likely differ from the demonstration program in several respects, which could lead to larger or smaller impacts than our estimates. Moreover, if the participating agencies are not representative of agencies nationally, the demonstration estimates may underestimate or overestimate the reductions in visits that would occur under a national prospective payment system.

Despite the limitations of these preliminary estimates, the conclusion is clear: prospective payment seems to be a promising alternative to the present payment system. If our future analyses confirm our belief that the impacts are not overstated because of underreporting, and no adverse effects on the quality of care or access to care are observed, prospective payment may be a viable mechanism for Medicare to reverse the pattern of large increases in home health care expenditures.

I. THE PER-EPISODE HOME HEALTH DEMONSTRATION AND EVALUATION

The Health Care Financing Administration's (HCFA's) Per-Episode Home Health Prospective Payment Demonstration tests the extent to which prospective payment for Medicare home health services increases efficiency in the provision of services. Such efficiency is meant to reduce public expenditures, while maintaining access to and quality of care. Per-episode payment encourages participating agencies to reduce the number of visits per episode and cost per visit, to generate savings. These incentives differ greatly from those found in the current system of cost-based reimbursement, which provides no reward for efficient care delivery.

This report presents the findings of an analysis of demonstration impacts for the early months of the demonstration on the provision of Medicare home health services during the period covered by the per-episode payment. These impacts are critical because a reduction in home health visits during this period is the chief means by which agencies may increase efficiency and earn profits. In addition, without an impact on such visits, we would not expect an impact on access to or quality of care.

This chapter provides an overview of the history of the Medicare home health benefit and the Medicare-certified home health industry, followed by a description of the Per-Episode Home Health Prospective Payment Demonstration. The reader who is familiar with Medicare home health may wish to begin with Section B, which describes the demonstration. Section C provides an overview of the research issues and general approach taken to conducting the evaluation.   Section D describes the hypothesized effects of per-episode payment on the provision of home health services.

A. THE MEDICARE HOME HEALTH BENEFIT

Congress established the Medicare home health care benefit in 1965, when the original Medicare program was created. Home health benefits were included to offer beneficiaries with acute conditions a less intensive and less expensive alternative to inpatient hospital care. At different times since the inception of the Medicare program, the home health benefit has been modified, partly to increase access to care.

Currently, the Medicare home health benefit covers home health services under Parts A and B; neither a deductible nor coinsurance applies. To be eligible for home health benefits, the beneficiary must (1) have Medicare coverage; (2) be homebound; (3) be under the care of a physician; and (4) need skilled nursing, physical therapy, or speech therapy on a part-time or intermittent basis.[1]

HCFA administers the Medicare home health benefit through fiscal intermediaries (FIs), each serving a defined geographic region of the country. In addition to serving as communication links between HCFA and the agencies, FIs also review claims to limit inappropriate use of services, determine reasonable costs, and administer payments to home health agencies.

Outside the prospective payment demonstration, Medicare reimburses agencies for the reasonable costs incurred to provide care. Since July 1987, an agency's per-visit costs have been judged reasonable as long as they do not exceed 112 percent of the mean cost incurred by all agencies (for the agency's mix of visits) in the same geographic area. Agencies incurring aggregate

---

[1]Skilled nursing services are covered as long as (1) a physician has ordered them, (2) the services are required on a part-time or intermittent basis, (3) the services require the skills of a registered nurse (or of a licensed practical nurse or licensed vocational nurse under a registered nurse's supervision), and (4) the services are reasonable and necessary to treat an illness or injury. Physical therapy and speech therapy are covered if a physician's assessment recommends it. Beneficiaries who need only occupational therapy are entitled to benefits only if they have established a prior need for skilled nursing care, speech therapy, or physical therapy in the current or prior certification period (see Teplitsky and Janson 1985-1992, p. VII.23, Section 204.4).

2

costs that exceed these limits are reimbursed only up to the limits (known as the Section 223 Limits). The cost limits were frozen for reporting periods that began July 1, 1994, through June 30, 1996.

Expenditures for home health care represent a material proportion of all Medicare expenditures (about nine percent in fiscal year 1994), and these expenditures have been growing rapidly in recent years (Health Care Financing Administration 1996). Spending for Medicare home health services has grown at least 20 percent a year since 1989, the year in which coverage was broadened as part of the settlement of a lawsuit brought against HCFA. After a 53 percent spike in annual growth in 1990, however, the rate of growth has declined (ProPAC 1996). Little of the growth is due to increases in cost per visit; rather, it is due to increases in the number of beneficiaries receiving Medicare home health care and the number of visits per beneficiary.

Since the program's inception, the number of Medicare-certified agencies has more than quadrupled. In 1995, there were roughly 8,700 Medicare-certified home care agencies (ProPAC 1996). Administratively, home health agencies have different ownership and auspices. They can be freestanding for-profit, freestanding nonprofit, affiliated with a facility (such as a hospital or skilled nursing facility), or operated by a government entity. Most of the recent growth in the number of Medicare-certified agencies has been in the number of hospital-based and freestanding for-profit agencies (ProPAC 1996). The distribution of ownership/auspices varies considerably by region of the country. Government-operated and private nonprofit agencies dominate the Northeast. Freestanding, for-profit agencies are pervasive in the South and West and even dominate the markets in some states.

Similarly, the number of Medicare-covered visits per episode and the length of episodes vary widely across regions. For example, among beneficiaries admitted to home health in 1990 and 1991, the mean number of approved visits in an episode of home health care was 47, and the mean episode

length was 94 days. However, the mean number of visits per episode varied from 28 in the Pacific region to 95 in the East South Central region, and the mean episode length varied from 60 days in the Pacific region to 180 in the East South Central region (Schore 1995). In 1994, the mean visits per beneficiary served were 66 nationally but varied from 45 in the Pacific region to 106 in the East South Central region (Health Care Financing Administration 1996).

The dramatic growth of home health as a proportion of total Medicare spending, combined with striking regional variation in its use and the explosive growth of the home health industry, prompted Congress to legislate changes to the Medicare home health benefit as part of the Balanced Budget Act of 1997. The act mandates the implementation of per-episode prospective payment for Medicare home health by 1999. Other changes to the home health benefit under the Balanced Budget Act include:

- A new algorithm for determining the maximum payment for an agency that is based on annual per-beneficiary costs or per-visit costs in a base year (whichever is lower)

- Elimination of coverage for blood drawing when it is the only home health service required

- Redefinition of "part-time" and "intermittent" care

- Redefinition of the payment basis from the location of the agency to the location of the patient

- Requirement of additional billing information (identifier for admitting physician and visit length)

B. THE PER-EPISODE DEMONSTRATION

The current Medicare home health payment system reimburses agencies for allowable costs up to a limit based on 105 percent of the median national cost. Because there is no mechanism for home health agencies to realize savings beyond costs, this system provides no incentive for producing

services efficiently and, in effect, subsidizes inefficient providers. Per-episode prospective payment is meant to increase efficiency, using the opportunity to generate savings as the primary incentive.

Ninety-one Medicare-certified home health agencies in five states--California, Florida, Illinois, Massachusetts, and Texas--enrolled in the three-year per-episode demonstration.[2] Forty-seven of them were randomly assigned to the treatment group to receive per-episode payment. The remaining 44 were assigned to the control group to continue under cost reimbursement. The first agencies in the treatment group began implementing prospective payment in June 1995; the latest entrant began in January 1996. Each agency started as its fiscal year began. Demonstration operations will continue through December 1998.

Mathematica Policy Research, Inc. (MPR) is the evaluation contractor responsible for assessing the impacts of the demonstration and its implementation. Several other organizations are participating in the demonstration. Abt Associates, Inc., is the implementation contractor responsible for recruiting demonstration agencies, monitoring the status of demonstration operations, and calculating certain statistics needed for agency payment., Palmetto Government Benefits Administrator (PGBA) is the FI responsible for review of claims and agency payment. The Center for Health Policy Research (CHPR) at the University of Colorado is the quality assurance contractor responsible for designing and implementing a quality assurance system for the demonstration agencies.

---

[2]Reflecting the United States more generally, considerable variation existed in the use of Medicare home health across the five demonstration states. In 1994, the mean numbers of visits provided per beneficiary using home health were: California, 46; Illinois, 52; Florida, 76; Massachusetts, 87; and Texas, 97 (Health Care Financing Administration 1996).

1. Demonstration Payment and Incentives

HCFA developed the Home Health Prospective Payment Demonstration to assess whether the profit motive can increase the efficiency of providing Medicare home health care and thereby reduce public expenditures, without sacrificing access to care or the quality of care. Phase I of the demonstration, which tested per-visit prospective rate setting, provided agencies an opportunity to generate profits (and avoid losses) by reducing per-visit costs.[3] The evaluation found that while agencies would make profits under this payment system, in order for the Medicare program to contain costs, Medicare must contain the volume of services. The current phase of the demonstration, Phase II, tests per-episode prospective payment. Under per-episode payment, agencies may earn profits by reducing the number of visits, as well as by reducing per-visit costs.[4]

a. Payment

Agencies selected for the treatment group receive a lump-sum payment for the first 120 days of home health care, regardless of number of visits provided or their cost.[5] The agencies are thus "at risk" for the costs of care incurred during this period. Those agencies that can provide care for less

---

[3] The per-visit demonstration was implemented in the same five states; however, most of the agencies participating in the per-episode demonstration did not participate in the per-visit demonstration. (Only agencies in the per-visit control group were eligible.) For details on the per-visit demonstration results, see Brown et al. (1995).

[4] Strictly speaking, only for-profit agencies earn profits; nonprofit agencies generate surpluses. However, for brevity, we use the term "profits" in this report to refer to surpluses generated by nonprofit agencies, as well as profits earned by for-profit agencies.

[5] Durable medical equipment, nonroutine medical supplies, and Part B ambulatory home health services continue to be reimbursed at cost throughout the demonstration. In addition, if an agency did not provide one or more of the six Medicare services during the base year but begins to do so during the demonstration, those visits are also reimbursed at cost during the demonstration, as are the costs of care for which Medicare is a secondary payer.

than the fixed (per-episode) rate will generate profits, whereas those whose costs exceed the fixed rate will incur losses.

For each visit beyond 120 days (referred to as outlier visits), treatment agencies receive a fixed payment rate that varies by the type of visit. In the demonstration, a treatment agency is also paid on a per-visit basis for visits made to patients admitted before the agency began demonstration operations ("phase-in" visits) and to those admitted within 120 days of the end of demonstration operations in that agency ("phase-out" visits). Agencies that can provide an outlier, phase-in, or phase-out visit for less than the fixed (per-visit) rate can also generate profits.

In the demonstration, home health episodes are defined by gaps in Medicare-covered home health care of at least 45 days. Only after the 120-day risk period and a 45-day gap in services can an agency receive a new per-episode payment for a given Medicare beneficiary.

b. Rate Setting

Prospective per-episode rates are based on an agency's costs and episode profile in the fiscal year preceding its entry into the demonstration (the base year), adjusted for inflation and changes in case mix in each evaluation year.[6] The episode profile is the average number of visits provided by the agency during an episode, calculated for each of the six types of visits covered by Medicare. Payment for outlier, phase-in, and phase-out visits are also based on the agency's base-year per-visit costs (adjusted for inflation).[7] HCFA's market basket is used to adjust both the per-visit and per-episode rates for inflation.

---

[6]For more information on payment rates under the demonstration, see Phillips et al. (1995).

[7]Because complete data for episode profiles and settled cost reports are not available for a given year until some months after that year is over, the initial lump-sum and per-visit rates used in the demonstration were preliminary and were revised as final base-year data became available.

The case-mix adjuster classifies each patient into one of 18 groups on the basis of 12 variables that describe the patient's characteristics. From this information, an aggregate case-mix index is created for each agency. At the end of each year of the demonstration, an agency's case-mix index for that year is compared with its case-mix index in the base quarter (the last quarter of the base year). If the agency's case mix differs, its aggregate payment is retrospectively adjusted (see Appendix C for further details).

c. Loss Sharing and Profit Sharing

To encourage agencies to participate in the demonstration, HCFA provided a loss-sharing arrangement. HCFA reimburses treatment agencies for 99 percent of losses in the first demonstration year, and for 98 and 97 percent of losses in the second and third demonstration years, respectively, subject to total payments being no greater than the Section 223 Limits.

To counteract the incentive to reduce the quality of care to generate profits, as well as to prevent agencies from realizing windfall profits at public expense, HCFA shares in profits above a specified threshold profit rate. If the total of a treatment agency's per-episode and per-visit prospective payments is greater than the costs incurred in rendering the services covered by these payments, profit greater than five percent of total allowable costs for these services is subject to profit sharing with HCFA. HCFA's share of profits is 25 percent if profits equal between 5 percent and 15 percent of total allowable costs. HCFA's share rises if profits exceed 15 percent (with the share of profits over 15 percent varying by demonstration year).

d. Incentives

Treatment agencies can reduce the cost of care rendered during the 120-day period by (1) reducing the number of visits provided during the risk period, (2) changing the visit mix to make less

costly visits a larger proportion of the total number, or (3) reducing per-visit costs (or some combination of these three). Reductions in the number of visits during the risk period could involve discharging patients earlier, thereby reducing the length of the episode or reducing the frequency of visits without reducing episode length. Reductions in the average number of visits could also be achieved by admitting a mix of patients needing less care, though this may also result in a lower payment. Reductions in per-visit costs could be achieved either by cutting direct costs (such as the length of a visit) or administrative costs (such as supervision). Alternatively, agencies might accept increases in per-visit costs to reduce the number of visits during the risk period. For example, agencies might hire wound care specialists (who command higher salaries), thereby reducing the number of visits, or they might use additional administrative resources to monitor the number of visits provided, thereby increasing per-visit costs. Per-visit costs might also increase if agencies perform in a single longer visit services that they previously provided in (and billed for as) two separate visits. In addition, as agencies reduce per-episode visits, they may experience a reduction in the direct-cost base over which their administrative costs must be spread, which may mean some loss in economies of scale. As a result, treatment agencies have an incentive to increase the number of outlier (and phase-in and phase-out) visits to help offset any volume reductions due to decreases in the number of visits during the risk period, as well as to increase the number of patients they serve.[8]

Profit motive is the prime incentive offered under the demonstration. While treatment agencies may incur losses, the generous loss-sharing provisions of the demonstration limit the incentive for agencies to alter their behavior to avoid losses, particularly in the first demonstration year. Thus,

---

[8]These issues will be addressed in future reports as data become available.

the demonstration's incentives rely heavily on the "carrot" of profits and relatively little on the "stick" of losses.

Agencies' responses to the incentives offered by the demonstration will depend on the priority each agency places on maximizing profits, relative to other goals. Nonprofit agencies, in particular, may view their primary mission as meeting the needs of the communities they serve. Consequently, they may be more reluctant than for-profit agencies to reduce visits during the risk period on the grounds that doing so would reduce care to those in need. The demonstration, however, does provide nonprofit agencies with an opportunity to generate profits that could then be used to develop programs that would benefit their community or to provide services to those in the community who cannot obtain them in other ways (such as through Medicaid, other public programs, or private purchase).[9]

Given the demonstration's emphasis on profit motive, we expect for-profit agencies to respond more aggressively than nonprofit agencies to the incentive of per-episode prospective rate setting. We also expect that hospital-based agencies may be less responsive than freestanding agencies to the opportunity to earn a profit under the demonstration. The former must respond to the hospital's need to discharge patients promptly and to "flow down" hospital administrative costs to the home health agency. Attention to the needs of the parent organization may also affect the behavior of other agencies that belong to a chain or other system of organizations.

---

[9] It is theoretically possible that nonprofit agencies might take advantage of the loss-sharing provisions to increase visits during the risk period, if they believe that Medicare has restricted the provision of needed care. Nonprofit agencies might treat the loss-sharing provisions as a source of community service funds, accessible with a small amount of funding (equal to one, two, or three percent of losses) from private sources.

2. Other Demonstration Procedures

a. Medical Review

For agencies in the treatment group, only limited medical review (known as "abbreviated" medical review) is performed by the demonstration FI for care delivered during the risk period. This review seeks to determine whether the patient met the coverage criteria for home health care and whether at least one visit that met these criteria was delivered. Only the admission bill is reviewed. As a condition of payment, the demonstration FI requires that the agency submit HCFA 485 and 486 forms (which contain information on the patient's health and eligibility status, as well as the home health plan of treatment) or clinical notes for admissions that coincide with an episode eligible for prospective payment. The medical review process is based on these materials.

All visits paid for under per-visit rate setting are subject to the usual focused medical review, under which a sample of claims is reviewed to ensure that each visit is medically reasonable and necessary. Medical review for control agencies continues under the current (nondemonstration) regulations. The only major difference is that control agencies are assigned to the demonstration FI. Since the demonstration FI's medical review procedures may differ in minor ways from those of other FIs, control agencies may be subject to somewhat different policies than those they are accustomed to.

b. Billing

Treatment agencies must submit an admission bill to the demonstration FI to initiate an episode of care. Treatment agencies are expected to submit interim bills for the rest of the risk period, although payment for visits is not predicated on their submission.[10] If there are outlier visits, the

---

[10] The interim bills are required to reimburse for supplies and to calculate costs for profit and loss (continued...)

agency must bill separately for those. When a patient is discharged, either during the risk or outlier period, agencies are to submit a discharge bill to terminate the episode. The FI will not initiate a new episode for a given patient unless a prior episode has been terminated. In addition, the FI checks that the 120-day risk period and a 45-day gap have elapsed before it initiates a new episode.

If a treatment admission claim is accepted (following abbreviated medical review), the per-episode payment is made as a lump sum.[11] While medical review is pending, subsequent episode bills are suspended. Initially, all episodes were subject to abbreviated medical review; in mid-1996, however, the proportion was reduced to 25 percent.[12] It was reduced because it took longer than expected to review the claims, delaying payments significantly; yet very few claims were denied. Since it is unrealistic to have 100 percent medical review in a national program, medical review was reduced.

Periodic interim payments (PIPs), which are intended to smooth cash flow for home health agencies, were originally discontinued for treatment agencies. However, a similar periodic payment system (called biweekly interim payments [BIPs]) was later reintroduced to meet the cash flow needs of some treatment agencies caused by delays in receiving per-episode payments.

---

[10](...continued)
sharing with HCFA. Interim bills also provide information on number of visits required for the evaluation.

[11]If the admission claim is denied, interim claims for that episode are suspended for 65 days to await appeal. If an appeal is filed, interim claims are suspended also until a decision is made on the appeal for the admission claim. When an admission claim is denied and an appeal is not filed within 65 days, or if the denial of the admission claim is upheld on appeal, suspended interim claims are released for possible payment under the agency's per-visit prospective payment rates.

[12]Abbreviated medical review was required for all episodes during most of the early months of the demonstration included in this analysis.

Control agencies continue to submit bills as under cost reimbursement and continue to be eligible for PIP. The FI bases PIP payments on the agency's average cost for each type of visit, while other FIs base PIP payment on overall agency average cost per visit. As a result, there may be minor differences in control agency PIP payments compared to what control agencies have experienced outside the demonstration.

c. Quality Assurance

All agencies participating in the demonstration (in both the treatment and control groups) are required to collect and submit patient-specific information to the demonstration quality assurance contractor. The quality assurance procedures follow a continuous quality improvement approach. Visiting staff from demonstration agencies are required to collect information (primarily on functional status and medical condition) at admission and at discharge, or 120 days after admission, whichever comes first. Similar information is also collected before admission to an inpatient facility (for a stay of 48 hours or more) and when the patient returns to home health care following such an inpatient stay. The quality assurance contractor uses this information to develop profiles describing patient outcomes at each agency. These profiles are provided to the demonstration agencies to help them improve the quality of care they provide.

C. COMPONENTS OF AND APPROACH TO THE EVALUATION

The evaluation seeks to answer two broad sets of policy questions: (1) How did home health agencies make decisions about participating in the demonstration, and how did they implement the demonstration? and (2) To what extent did per-episode prospective payment for home health affect the behavior of home health agencies and outcomes for their patients--that is, what was its impact? These areas of inquiry are interrelated. Impacts can be interpreted only in light of how per-episode

prospective payment was implemented by agencies participating in the demonstration and within the context of how the demonstration operated. It is critical to understand which strategies help produce more efficient home health care and how government can best shape policies to encourage such behavior by agencies with widely differing missions and characteristics. In addition, the nature and extent of program impacts must be determined in order to properly focus analysis of decisions and operations on agencies that were particularly successful (or unsuccessful) in improving efficiency. The integration of these two analyses is a key aspect of our evaluation. (For a detailed discussion, see Phillips et al. 1995.)

1. Analysis of Agency Decisions and Operations

The analysis of agency decisions and operations focuses primarily on five questions:

1. What factors explain the decision of home health agencies to participate in the demonstration?

2. Was the demonstration implemented as planned?

3. What strategies did agencies adopt to reduce visits per episode and per-visit costs?

4. What were the effects of these strategies on the process of care, and what implications do these effects have for access to care and the quality of care?

5. What factors explain the key decisions of home health agencies?

The evaluation took a case study approach to answering these questions. This approach included judgmentally selecting 67 of the agencies participating in the demonstration for site visits at the start and conclusion of the demonstration. The goal in selecting agencies for site visits was to obtain as much information as possible about various agency environments, decision-making processes, and responses to the demonstration. The site visit data are supplemented with program records and discussions with other demonstration actors.

The following were the key conclusions of the first report on agency decisions and operations (Phillips and Thompson 1997), which was based on site visits conducted during the first half of 1996 (and the supplemental material just described):

- Most agencies saw the demonstration as an opportunity to learn about operating under prospective payment with limited financial risk. Over half (including both for-profit and nonprofit agencies) saw generating a profit or surplus as another key objective.

- Most agencies characterized the home health environment in which they operated as highly competitive, although hospital-based agencies perceived competition as less intense (probably because they had "protected" referral streams). Nearly half of the agencies receiving site visits reported that there was an oversupply of agencies in the communities they served.

- More than 80 percent of agencies had Medicare managed care plans operating in the areas they served. This varied from 100 percent of California and Florida agencies to 70 percent of Texas agencies. Even though managed care was widespread, however, most demonstration agencies served few managed care enrollees.

- Agencies expected growth rates during the demonstration that were dramatically different from those that prevailed only a few years ago. About half expected to increase the number of patients seen, with anticipated growth rates averaging about nine percent a year. About 15 percent of agencies expected to shrink, with expected declines in caseloads averaging about 8 percent. Agencies cited a number of reasons for the expected decline in visits, including the growth of managed care and, for treatment agencies, the expected decline in visits per episode.

- Treatment and control agencies planned to reduce their costs per visit during the demonstration. The most common strategy planned was to reduce administrative costs; another approach was to increase use of technology.

- About half of treatment agencies planned to reduce per-episode costs. Strategies included reviewing utilization more intensely, rationalizing the process of care through care maps (or critical pathways), placing greater reliance on community services and family caregivers, and increasing use of telephone contact with patients.

A second report will investigate the implementation of demonstration procedures and responses to demonstration incentives as they evolved during the three-year demonstration. We will also

prepare a quantitative analysis of the factors associated with agency participation in the demonstration.

2. Analysis of Program Impacts

In this section, we provide an overview of the issues, data, and methods for the overall analysis of program impacts. The methods and data for this preliminary analysis of impacts on home health use are described in detail in Chapter II.

a. Research Issues

Many of the critical policy issues to be addressed in the evaluation pertain to program impacts-- the extent to which per-episode prospective payment for home health care affects the behavior of home health agencies and outcomes for their patients. Controlling public expenditures for home health care would be a primary objective of a national program of prospective payment for home health agencies. A key aim of the evaluation, therefore, is to measure the impact of per-episode prospective payment on per-episode service use, to determine the potential for savings. Because per-episode prospective payment may alter per-visit costs and the mix of visits rendered, as well as the number of home health visits, it will be necessary to identify the relative importance (to any expenditure reductions) of changes in the cost of producing a visit and in the number and types of visits rendered. It will also be important to assess whether effects on service use and agency behavior are likely to affect access to or quality of care, as well as the extent of any such effects. Through its potential effect on access to and quality of care, per-episode prospective payment may potentially shift care to nursing homes or hospitals, to programs that provide community-based services, or to informal caregivers (that is, family members and friends). The evaluation will also

identify the extent of such shifts and implications for the overall burden and cost of care borne by public programs (including Medicaid) and informal caregivers.

These policy issues suggest the following key research questions concerning demonstration impacts:

- What effect does per-episode prospective payment have on Medicare home health services received during the risk period, the outlier period, and overall?

- What effect does per-episode prospective payment have on per-visit costs for Medicare-certified home health agencies and on the volume and types of services provided by these agencies?

- What effect does per-episode prospective payment have on patient selection and retention and, thus, on access to care?

- What effect does per-episode prospective payment have on quality of care?

- What effect does per-episode prospective payment have on Medicare expenditures generally?

- What effect does per-episode prospective payment have on the use of and expenditures for non-Medicare-covered services, including the use of Medicaid services, other home and community-based services, and informal care?

- Do the effects of per-episode prospective payment vary with the characteristics of the patient or the agency?

b. Methods

Addressing the many complex issues involved in evaluating the impacts of per-episode prospective rate setting requires data from many sources, numerous samples, a comprehensive research design that allows observation of intermediate results, and sophisticated analytic procedures that maximize the available observations. We will estimate program impacts statistically, exploiting the randomized design (by comparing outcomes for treatment and control groups) and the panel nature of the data.

Our approaches to estimating impacts on different outcomes are similar in principle; they differ only in the particulars that concern the unit of analysis, the samples and data sources, and the statistical model. For each outcome measure, we will compare the experience of the treatment agencies (or their patients) with that of the control agencies (or their patients) during the demonstration period.

For most of the analyses, the episode will be the unit of analysis. The fact that payment will be fixed for an episode provides a compelling reason for using the episode as the analysis unit. While patients could be used as the unit, some will have multiple episodes. We will, however, use the patient-year as the unit of analysis for investigating impacts across episodes. Finally, we will use the agency as the unit of analysis for investigating impacts on agency outcomes, such as cost per visit and structural aspects of quality.

Outcome measures are drawn from secondary and primary sources. The primary sources include: Medicare standard analytic files and enrollment files, data collected by the demonstration quality assurance contractor, State Medicaid Research Files (SMRFs), Medicare cost reports, and demonstration Uniform Billing (UB-92) forms. MPR is conducting three primary data collection efforts. First, we are conducting telephone interviews with a sample of patients three and eight months after their admission to home health to obtain data on their health and functional status, satisfaction with home health care, and their use of non-Medicare services. Second, we have asked demonstration agency nurses and aides to provide data from a single day on the length of their Medicare home health visits. Finally, we are asking agencies to complete an annual self-administered survey, which will provide information on agency characteristics and procedures, including structural measures for the analysis of care quality.

Control variables will be derived from these same sources, as well as from the Section 223 limit files, the Provider of Services File, and the Area Resource File (ARF).

We will use regression analysis or a related multivariate statistical technique to control for exogenous differences that may exist between treatment and control groups despite agencies' random assignment to treatment or control status. For cross-sectional data sets, we will use ordinary least squares regression for continuous dependent variables and logit analysis for binary dependent variables. For analyses with agencies as the unit of observation and data for multiple years, we will use fixed- and random-effects regression models. Observations are weighted in the analyses so that each agency has equal representation. Standard errors are calculated using special software that takes into account the effects of clustering and weighting.

D. GUIDE TO THE REST OF THIS REPORT

The second chapter of this report describes the data sources, samples, and specific analytic approaches used in this analysis of impacts on home health service use for the first year (approximately) of the demonstration. Chapter III presents our hypotheses and describes findings, while Chapter IV presents our conclusions concerning this analysis.

II. DATA AND METHODOLOGY

We estimate program impacts on measures of home health care utilization claims data. Regression models are used to obtain the estimates, to control for possible preexisting differences between treatment and control agencies, or differences in their patients, that could be confounded with the impacts of prospective payment.

A. DATA

The measures of demonstration home health use in this analysis were constructed from Medicare claims data for patients admitted to demonstration agencies between the agency's demonstration start date and August 31, 1996--the first 8 to 15 months of the demonstration. The episodes were identified and constructed from demonstration UB-92 bill record files obtained from the demonstration FI. Data from HCFA's standard analytic files were used to construct home health service use outcome measures pertaining to the 120-day period for each episode identified from the UB-92 files.

To construct measures to control for possible differences between the treatment and control groups in patient, agency, and area characteristics, we drew on several data sources. Standard analytic file data were used to construct measures of use of Medicare services by demonstration patients prior to admission and measures of home health services rendered by demonstration agencies in the predemonstration period. Other measures of patient characteristics were drawn from HCFA's Enrollment Database (EDB) file. We also drew on demonstration data collected for case-mix adjustment on each demonstration patient at admission and for each agency's patients in the quarter preceding the agency's enrollment in the demonstration (base quarter). Finally, we drew on agency cost reports for the base year and on the ARF.

Some of the variables used to control for agency characteristics also define subgroups (for example, proprietary and nonprofit agencies) for which we will investigate differences in demonstration impacts.

From these varied sources, we constructed an analysis file in which each observation contains data on (1) the patient's use of Medicare-covered services during the home health episode; (2) the patient's characteristics at admission to home health; (3) the patient's Medicare service use before the home health admission; (4) the predemonstration characteristics of the home health agency providing care, including characteristics of its patient mix; and (5) the characteristics of the area in which the agency is located.

1. Identifying Episodes

For patients of both treatment and control agencies, data from UB-92 bill record files obtained from the demonstration FI, PGBA, were used to identify home health episodes as defined by demonstration rules. Beginning with each agency's enrollment in the demonstration, we scanned the UB-92 files to identify the first admission for each individual and all that person's subsequent bill records. To create an episode, we combined all records for an individual for 120 days following the first admission and any bills for care after 120 days until we observed a gap of at least 45 days in billing dates. This procedure was followed regardless of whether the agency discharged and readmitted a patient during the 165 (120 + 45) days. If we observed additional home health bills after 165 days, we created a second episode for that individual, and so on for any subsequent episodes beginning through August 31, 1996.

2. Medicare Claims for Outcomes Variables

We matched the episodes identified with UB-92 bill record data to HCFA's Medicare standard analytic files. Matches were identified for 57,261 episodes; only 493 episodes (less than one percent) did not match.

Using standard analytic file data, we constructed outcome measures for demonstration home health service use. These outcome measures are listed in Table II.1 and fall into three categories: (1) number of home health visits, overall and by type; (2) types of services; and (3) episode length. Only Part A home health data were extracted for this analysis since any home health care provided under Part B is not eligible for per-episode payment. In addition, we extracted claims only for the demonstration agency admitting the patient for a given episode. Impacts on services rendered by other home health agencies during the 120-day period will be addressed in a forthcoming report.

We used the standard analytic file data rather than the UB-92 data to construct the outcome measures because HCFA adjusts the standard analytic file data, but not the UB-92 data, for any voided or amended bills. While we could have mimicked HCFA's adjustment algorithms, we could not be sure that we would replicate all their adjustments perfectly.

Medicare claims were extracted from the standard analytic files in May 1997. Since claims are generally included in the standard analytic file within four months after the service was rendered, we should have nearly complete data on home health services received through December 1996.[1]

3. Control Variables

As Table II.2 indicates, we used several types of control variables in this analysis. Patient characteristics at the start of an episode and patient Medicare service use in the six months preceding

---

[1]This is why we chose the August 31 cutoff date, since it allows for complete data to be extracted for the entire at-risk period.

TABLE II.1

OUTCOME VARIABLES DESCRIBING MEDICARE SERVICE USE
DURING THE 120-DAY AT-RISK PERIOD
(Part A Services Only)

Number of Visits
   Total visits
   Skilled nursing visits
   Home health aide visits
   Physical therapy visits
   Occupational therapy visits
   Medical social worker visits

Type of Services
   Probability of any given service
   Proportion of skilled nursing visits
   Proportion of home health aide visits
   Proportion of medical social worker visits
   Number of days until the first medical social worker visit

Episode Length
   Length of an episode within at-risk period
   Probability of an episode lasting beyond the at-risk period

TABLE II.2

CONTROL VARIABLES FOR MULTIVARIATE ANALYSIS, BY SOURCE

Episode Level		Agency Level		Area Level
Patient Characteristics at Episode Start (Medicare Enrollment Database; UB-92 remarks)	Medicare Service Use in Year Preceding Episode (Medicare Standard Analytic Files)	Base-Quarter Patient Service Use (Abt Base Quarter Case-Mix File)	Agency Characteristics (Base-Year Cost Reports and Abt Enrollment File)	Area Characteristics (Area Resource File)
Age	Length of pre-home-health inpatient stay	Agency practice pattern index	Chain member	Physicians per 10,000 (1994)
Gender			Hospital-based	Nursing home beds per 100 elderly residents (1991)
Race	Whether in skilled nursing facility within 14 days before episode start		Proprietary	Hospital occupancy rate (1993)
Original reason for entitlement			Small agency (less than 30,000 visits in base year)	
Whether has cancer	Number of home health visits in 6 months prior to episode start		State	
Whether has diabetes			Rural	
Whether has decubiti	Total Part A Medicare reimbursement in 6 months prior to episode start			
Needs complex wound care				
Limitations in activities of daily living				
Whether admitted to home health from hospital				

the episode were used to control for differences between the patients of treatment and control agencies. Agency and area characteristics were used to control for differences between the treatment and control agencies that might influence use of home health services.

Patient Characteristics. Patient characteristic control variables were used in the analysis to account for possible differences in patient mix between treatment and control agencies. We expect that individuals who are more severely ill, have diagnoses requiring greater amounts of care, and have greater limitations in daily activities will require more Medicare home health services.

We obtained data on patient characteristics at the start of the home health episode from three sources: (1) UB-92 patient characteristic data collected for the demonstration's case-mix adjuster, (2) Medicare enrollment databases, and (3) Medicare standard analytic files. In the remarks field for the first UB-92 bill following a demonstration admission, both treatment and control agencies were required to submit the information on patient characteristics needed for the 18-category Home Health Utilization Group (HHUG) case-mix adjuster. The characteristics include measures of impairment in Activities of Daily Living (ADL) and whether the patient has certain medical conditions (cancer, diabetes, decubiti) and care needs (complex wound care). Medicare enrollment files provide us with basic patient demographic information, including the patient's age (at the start of home health episode), gender, race, and disability status (from the original reason for Medicare qualification). From the standard analytic files, we constructed measures of Medicare service use to capture the patients' severity of illness, including measures of recent acute illness (whether admitted from hospital, length of prior hospital stay) and longer-term home health use (six months prior to admission). For the latter, we used the mean value for beneficiaries between 65.5 and 66 years old as a proxy measure for beneficiaries less than 65.5 years old at home health admission,

since beneficiaries under age 65.5 would not have been eligible for Medicare service for a full six months.

Agency and Area Baseline Characteristics. Agency characteristics are used as control variables because different types of agencies may have different goals and different cost and management structures, which could affect the home health care they render. In addition, certain types of agencies may serve a mix of patients requiring more (or less) care than the patients of other agencies. For example, proprietary and nonprofit agencies might have different preexisting practice patterns with respect to the number of visits rendered per episode, and hospital-based agencies might serve somewhat more-acutely ill patients than freestanding agencies.

Agency characteristics are also used to define subgroups because agencies with different goals, cost and management structures, and practice patterns could respond differently to the incentives of the demonstration. For example, proprietary agencies may have a stronger interest in revenue surpluses (profits) than nonprofit agencies and therefore reduce visits by a greater margin.

To define agency size, we used the total number of visits in the agency's base year. We defined small agencies as those that rendered fewer than 30,000 visits per year (approximately 25 percent of the sample). Large agencies are those that rendered 30,000 visits or more in the base year.

Data on agencies' structural characteristics were obtained from base-year Medicare cost reports and from the demonstration implementation contractor, Abt Associates. Base-year Medicare cost reports provided information on the agencies' base-year characteristics, including for-profit status, affiliation, and size (as measured by total number of visits rendered). Measures constructed during demonstration recruitment by Abt Associates that were used included factors such as whether the agency was a member of a chain or was located in a rural area according to the census definition.

We also developed a control variable measuring each agency's predemonstration practice pattern for the 120-day period. This practice pattern variable is an index of the average number of visits per episode provided by an agency in the first 120 days of base-quarter episodes relative to the average number provided by other demonstration agencies. The index accounts for differences across agencies in the average number of visits of each type and the characteristics of the patients.[2] A value greater than one indicates that, controlling for differences in case mix, an agency provided more visits during the 120-day period than did other demonstration agencies during the quarter preceding the demonstration.

Area-level characteristics that might influence the care an agency renders were also controlled for. For example, in areas where the number of nursing home beds is limited (relative to demand), hospitals may discharge to home health care some patients who otherwise would be discharged to nursing home care. We obtained area characteristics from the ARF, including physicians per 10,000 residents, nursing home beds per 100 elderly residents, and hospital occupancy rates.

---

[2]The specific construction is as follows. Let subscript i refer to the service type and subscript j refer to case-mix cell. Using the case-mix adjuster (developed by Abt Associates) an agency's patients into one of 18 case-mix cells (j=1,..,18). Within each case-mix cell, we multiply the average visits of each type for a given agency (n_{ij}) by its national cost limit in the base year (w_i), and then sum across each of the visit types. This sum essentially reflects a weighted count of the average visits for an agency within a case-mix cell. Second, we use these weighted counts to construct the ratio of the agency's average number of visits received by patients in the case-mix cell to the average among all agencies for this cell ($\Sigma w_i N_{ij}$). Finally, for each agency, we arrive at the practice pattern index by summing across the 18 case-mix ratios, weighting each ratio by the agency's proportion of episodes in the case-mix cell during the base quarter (p_j). Thus, for each agency, the index practice pattern is given by:

$$\sum_j p_j \left(\frac{\sum_i w_i n_{ij}}{\sum_i w_i N_{ij}} \right)$$

4. Analysis Sample

Our analysis sample contains 50,691 episodes drawn from 85 of the participating agencies (44 treatments and 41 controls).[3] From the initial set of 91 agencies (47 treatment and 44 controls), three control agencies were dropped because they withdrew from the demonstration after participating for only a few months, one treatment group agency was dropped because it had no admissions, and two other treatment agencies were dropped due to serious data problems.[4] Excluding all episodes from these agencies, plus individual cases not eligible for the demonstration or having missing data on key control variables, resulted in the loss of 6,570 of the 57,261 episodes initially constructed.

The three agencies that dropped out of the demonstration early either were purchased by another agency or were merged with another agency, and the new ownership did not want to be part of the demonstration. Loss of these agencies accounted for 1,613 episodes. The two dropped because of data problems had incomplete claims for an unknown number of episodes. Exclusion of these agencies caused 796 episodes to be lost from the analysis.

In addition to the loss of observations from dropping agencies, we dropped a number of individual episodes for three reasons. We dropped 1,114 episodes because we learned from the HCFA enrollment files that these patients were enrolled in a Medicare health maintenance organization (HMO) at the time of their home health admission and were not eligible to participate in the demonstration. The admitting demonstration agencies were probably not aware that these patients were HMO members when they admitted them and submitted a claim to the demonstration FI in error. We excluded another 1,267 episodes because Medicare was a secondary payer.

---

[3]These episodes correspond to 49,391 different individuals, less than 3 percent of which had more than one episode in our sample.

[4]See Appendix A for an examination of the data quality.

Agencies were not entitled to a per-episode payment for these episodes. Treatment and control agencies are both reimbursed at cost during the demonstration for care for which Medicare is a secondary payer. Finally, we excluded 1,735 episodes that were missing the patient characteristic data from the remarks section of the UB-92. These data had been inadvertently erased from the UB-92s because of an error in the software used by PGBA, the demonstration FI. PGBA is working to restore the data but could not complete the task in time for this report.

5. Summary Statistics for Control Variables

When the number of units randomized is not large, as is the case here (91 agencies), random assignment cannot be relied upon to yield treatment and control groups that are identical at baseline. Furthermore, as noted above, three agencies dropped out of the demonstration and two were excluded from the analysis, which also could create differences between the two groups. Indeed, the two groups are not equivalent here; they differ on a number of observed characteristics. Therefore, we cannot rely on simple treatment-control differences to produce unbiased estimates of program impacts; we must control for baseline differences in the treatment and control groups.

Table II.3 displays the treatment and control group means for the explanatory variables in our regression models. (Following the methodology described in Section B, these means have been constructed using sample weights that give each agency equal representation.)[5]

---

[5]The significance levels for the tests of equality between treatment and control group means in Table II.3 do not account for design effects due to the clustering. For our purposes here (but not in the main analysis described in Chapter IV), there is no need to account for these effects because we are interested only in differences within this sample, not the population of all agencies. We do account for the design effects associated with our use of sample weights, however. See Section B of this chapter for a complete discussion of the use of weighting and clustering in our analysis.

TABLE II.3

WEIGHTED MEANS FOR EXPLANATORY VARIABLES BY TREATMENT
STATUS, AND TESTS FOR DIFFERENCES IN TREATMENT
AND CONTROL GROUP MEANS

Explanatory Variable	Treatments	Controls
Demographic Measures (Percentage)		
Age Was Original Reason for Medicare Eligibility	83.3	81.3***
Age: Less than 65 Years Old	8.0	9.0*
Age: 75 to 84 Years Old	39.7	40.0
Age: Older than 84 Years	23.5	21.6***
Race: White	81.6	81.0
Female	63.1	64.4
Medical Conditions and Care (Percentage)		
Cancer	12.7	12.8
Diabetes	21.7	21.7
Cardiovascular Accident	15.3	14.7
Decubitus Ulcer: Stage 3 or 4	4.6	3.7***
Need Complex Wound Care	7.0	7.0
Limitations in Activities of Daily Living (Percentage)		
Bathing	71.8	72.8
Eating	28.0	30.1
Dressing	60.5	64.6***
Toileting	38.0	40.7***
Transferring	49.7	52.2 ***
Patient Prior Service Use Measures		
Admitted to Home Health from Hospital (Percentage)	35.2	37.8***
Length of Hospital Stay Ending Within 14 Days Before Home Health Start	3.6	4.2***
Whether Had a SNF Stay Ending Within 14 Days Before Home Health Start (Percentage)	17.7	15.3***

TABLE II.3 *(continued)*

Explanatory Variable	Treatments	Controls
Total Part A (Plus Part B Inpatient/Skilled Nursing Facility/Home Health/Hospice) Reimbursement During 6 Months Before Home Health Start	$11,415	$11,341
Number of Part A Home Health Visits During 6 Months Before Home Health Start	12.2	13.9 **
Had Medicare for Less than 6 Months at Episode Start (Percentage)	1.5	1.4
Agency Characteristics Measures (Percentage)		
Proprietary	47.7	51.2***
Hospital-Based	19.1	14.6**
Chain	38.6	26.8*
Small Agency (Less Than 30,000 Visits in Base Year)	34.1	19.5 *
Agency Practice Pattern: Index of Average Visits[a]	0.94	1.10***
Area Characteristics Measures		
Florida (Percentage)	6.8	9.8**
Illinois (Percentage)	13.6	22.0***
Massachusetts (Percentage)	18.2	7.3**
Texas (Percentage)	36.3	39.0**
Urban (Percentage)	84.1	85.3
Number of Physicians per 10,000 Persons (1994)	22.0	21.5***
Number of Nursing Home Beds per 100 Persons over 65 (1991)	5.2	5.2
Hospital Occupancy Rate (1993) (Percentage)	62.4	60.8**
Sample Size	**25,561**	**25,130**

[a]An index of the case-mix adjusted average visits received by an agency's patients in the first 120-days of base quarter episode, relative to the average across all agencies.

*Difference in means are significantly different from zero at the .10 level, two-tailed test.
**Difference in means are significantly different from zero at the .05 level, two-tailed test.
***Difference in means are significantly different from zero at the .01 level, two-tailed test.

With the large sample of episodes that we have available, we have the statistical power to detect very small differences between the treatment and control groups at baseline.[6] We therefore expect that many differences in the explanatory variables between the treatment and control agencies will be statistically significant even when the magnitude of the difference is not large.

Demographic Measures. Some of the demographic characteristics of patients admitted to treatment and control agencies differed significantly, but the magnitudes of the differences are not large. Relative to control agencies, treatment agencies served a slightly higher proportion of patients age 85 and older and about two percent more patients originally entitled to Medicare because of age.

Medical Conditions and Care. There was one material difference between treatment and control agencies in the means for measures of medical conditions and care needs drawn from the case-mix data in the remarks field of the UB-92. Treatment agencies were significantly more likely to serve patients who had a decubitus ulcer, with a 24 percent increase in the treatment mean, relative to the control mean.

Limitations in Activities of Daily Living. Relative to control agencies, treatment agencies served patients who were less likely to be impaired in ADLs, but the magnitudes of the differences are quite small. For three of the five ADL tasks for which we have measures (dressing, toileting, and transferring), treatment agencies had significantly lower sample means than control agencies. All these treatment-control differences are less than 10 percent of the control group mean.

Patient Prior Service Use Measures. Patients served by treatment and control agencies differed somewhat in their use of medical services prior to entering a home health episode. Treatment agencies served patients who were more likely to have been recently discharged from a

---

[6]For a binary indicator with a mean of 50 percent, for example, the minimum detectable difference is 1.3 percentage points, using a two-tailed test with a 95 percent confidence interval, at 80 percent power.

skilled nursing facility and less likely to have been admitted after release from a hospital. These differences offset one another; each is roughly two-and-a-half percentage points. In addition, prior hospital stays were shorter for patients in the treatment group than in the control group. For the treatment group, the mean length for a prior hospital stay was about 14 percent shorter than for the control group. Patients of treatment agencies also received 1.6 fewer home health visits in the six months prior to the demonstration episode, 12 percent less than the control group.

Agency Characteristics Measures. There are several significant differences in the characteristics of treatment and control agencies, and a few of these differences are large. While treatment agencies are significantly less likely than control agencies to be proprietary, this difference is not large. On the other hand, treatment agencies are significantly more likely to be (1) hospital-based, with a difference of 31 percent from the control group mean; (2) affiliated with a chain, with a difference of 44 percent from the control group mean; and (3) small, with a difference of 15 percentage points, representing an increase of about 75 percent from the control group mean of 20 percent.

We also observe a material and statistically significant preexisting treatment-control difference in practice patterns. On average, the practice pattern index for the treatment agencies is about 15 percent lower for treatment group than for control group agencies.[7]

Area Characteristics Measures. Significant differences also exist between treatment and control agencies in area characteristics. The distribution across states differs for the treatment and control agencies, with the treatment group particularly overrepresented in Massachusetts and underrepresented in Illinois relative to the control group. There are also significant differences in

---

[7]Evaluated at the mean visits per episode for all agencies in the base year (43.7), this pre-existing difference in practice patterns corresponds to an additional seven visits provided by the average treatment agency compared to the average control.

the rate of physicians per 10,000 residents and the hospital occupancy rate; however, both of these differences are small, representing less than a three percent difference relative to the control group mean.

In summary, while the preexisting treatment-control differences in patient characteristics are minimal, there are several large differences in agency and area (state) characteristics. If these preexisting differences affect the provision of home health services, we would incorrectly estimate the effects of per-episode payment if we were to compare simple treatment and control group means. In fact, the large and extensive differences in agency and area characteristics suggest that a simple comparison of treatment and control group means could be very misleading. Because regression models allow us to control for these preexisting differences, they may provide more accurate estimates of the effect of the demonstration payment method.

B. METHODS

To assess the impacts of prospective payment, we estimate multivariate models for all episodes occurring in treatment and control agencies during the initial period of the demonstration. The models predict the use of home health services as a function of whether a patient is being treated by a treatment or control agency, the patient's personal characteristics, and the characteristics of the agency and the area. The estimated impact of the payment method is given by the difference in regression-adjusted means between treatment and control agencies. We use regression-adjusted means because they improve the precision of the estimated effects and control for the preexisting treatment-control differences discussed earlier in the chapter.

We use three types of weighted regression models. For our main findings, we estimate impacts using ordinary least squares (OLS) regression when the dependent variable is continuous and logistic regression when the dependent variable is binary. As part of the sensitivity analysis, we also

estimate impacts using Tobit regression models when the dependent variable is continuous but has been censored at a discrete value (for example, number of physical therapy visits). The main regression models use weighted data so that each agency is given equal representation in the analysis; however, we also examine the robustness of our results with respect to alternative sample weights. All analyses use standard errors that take account of the effects of sample clustering and weighting, as described below.

1. Statistical Models for Estimating Overall Impacts

When the outcome that we investigate is continuous (for example, number of visits, episode length), the basic model that we use to estimate the overall impacts of prospective payment is:

(1) $Y = \alpha + X\beta + \delta T + \epsilon,$

where

Y is a continuous outcome variable for episodes taking place during the initial period of the demonstration

X is a vector of control variables

T is a binary variable for treatment status that equals 1 for episodes rendered by treatment agencies and 0 for episodes rendered by control agencies[8]

α is the intercept term

β is the vector of regression coefficients on the control variables

δ is a parameter that measures the impact of prospective rate setting on the outcome Y

ϵ is a random disturbance term assumed to have a mean of zero (conditional on X and T) that reflects all the unobserved factors affecting Y

---

[8]The omitted (reference) binary variable equals 1 for episodes rendered by control agencies and 0 for episodes rendered by treatment agencies.

The coefficient δ on the variable T measures the effect of the demonstration payment method on the (continuous) outcome of interest and is tested to determine whether it is significantly different from zero. For the main analysis, we use OLS to estimate this equation; however, when the dependent variable is censored, we also investigate impacts using Tobit models (see Section E).

When the outcome variable that we investigate is binary (for example, whether a home health aide visit was rendered, whether an episode exceeds 120 days), the logit model is used to estimate demonstration impacts. The structure of the logit model is as follows:

$$(2) \quad Probability\ (Y{=}1) \;=\; \frac{1}{1 \;+\; e^{(\alpha \;+\; X\beta \;+\; \delta T)}},$$

where Y is the (binary) outcome variable and the remaining variables and parameters are defined as in equation (1).

Given the nonlinearity of the logit model, the estimated impact of the payment method is not measured directly by the coefficients δ on the variable for treatment status. To estimate the demonstration impact on the probability that $Y = 1$, we use the coefficient estimates from the model to generate two predicted probabilities for each observation: one assuming that the observation belongs to the treatment group ($T = 1$), and one assuming that it belongs to the control group ($T = 0$). The impact estimate is the average difference between these estimated probabilities. Because the statistical significance of δ determines whether the odds that $Y = 1$ are significantly different for the treatment and control groups, we use the p-value on this parameter to test our hypotheses about differences between the two groups.

Throughout the tables of results, we present the unadjusted mean of each outcome variable for the control group alongside the estimated impact, as a point of reference. This statistic provides for the outcome variable a reasonable estimate of the mean value that might be expected to occur in the

absence of the demonstration. We use this mean to assess the relative magnitude and importance of the estimated impact.

2. Estimating Subgroup Impacts

We investigate demonstration impacts for agency subgroups for two reasons. First, if impacts exist for the full sample, we are interested in whether they differ between subgroups of agencies defined by their profit status and other important characteristics. Second, even if we find no compelling evidence of overall impacts, it is important to determine whether they exist for selected subgroups.[9]

The subgroup analysis uses a regression model similar to equation (1). The only difference is that the subgroup model includes additional control variables formed by interacting treatment status T with a set of binary (subgroup) variables defined from agency characteristics. Specifically, for each of four agency characteristics--profit-status, pre-existing practice patterns, size, and auspice--we form an indicator variable which takes on the value of one if the agency belongs to the subgroup and zero otherwise.[10] The regression model is thus given by:

$$(3) \quad Y = \alpha + X_1\beta_1 + X_2\beta_2 + \delta T + \sum_{i=1}^{4} \gamma_i (x_{1i} * T) + \epsilon$$

where

X_1 is a vector of (four) subgroup variables, each of which takes on a value of one if the agency belongs to the subgroup and zero otherwise

---

[9]Future analyses will also investigate whether impacts vary with patient characteristics.

[10]As described in Chapter III, the pairs of subgroup variables defined from these characteristics are as follows: (1) for-profit status and non-profit status; (2) high-use practice patterns and low-use practice (reference) patterns; (3) small size and large size; and (4) freestanding and hospital-based. Since a (reference) variable in each pair must be dropped from the regression model for identification, only four variables--indicating for-profit status, high-use practice patterns, small size, and hospital-based auspice--actually enter the estimated equation.

X_2 includes all other control variables

and the remaining variables and parameters are defined as in equation (1).

To assess whether the impacts differ between particular subgroups, we examine the statistical significance of the coefficient (γ_i) on the corresponding interaction term. For example, suppose that the p-value for the coefficient on the interaction term between an agency's for-profit status and its treatment status is .04 (statistically significant from zero). This test indicates that the effect of prospective payment on that outcome measure differs significantly between for-profit and nonprofit agencies.

We also test whether the effect of prospective payment is significantly different from zero for each subgroup.[11] The impact for a subgroup is estimated by setting the indicator variable for the subgroup appropriately and then obtaining the predicted value of the outcome variable for each observation, first as if it were from a treatment agency ($T = 1$), and then as if from a control agency ($T = 0$). The mean difference between these two predicted values provides an unbiased estimate of the impact of prospective payment on the outcome Y for each subgroup. Thus, for the subgroup of agencies with characteristic j, the impact estimate is given by:[12]

$$(4) \quad \delta + \sum_{i \neq j} \gamma_i \overline{x_{1i}} + \gamma_j$$

---

[11]It may not be possible to identify small or moderate impacts for subgroups, since tests of statistical significance lose power as the sample size declines. The smaller the size of the subgroup (all else equal), the less likely we are to reject the hypothesis that the payment method has no effect, for any given true effect size.

[12]This expression assumes that the subgroup variable defined by characteristic j is not the (reference) variable dropped for identification. If it is the reference variable, the impact would be given by the first two terms in equation (4) only.

3. Hypothesis Tests for the Impact Estimates

For each outcome, a two-tailed t-statistic tests the null hypothesis that there is no difference between the regression-adjusted means for treatment and control agencies.[13] The associated p-value is used to determine whether the demonstration had a measurable impact. The p-value is based on estimated standard errors that account for the clustering of episodes within agencies and the use of sample weights. A p-value below 0.10 indicates rejection of the null hypothesis and provides significant statistical evidence that a demonstration impact exists. At this p-value, however, approximately 10 percent of independent tests will show, simply by chance, a statistically significant treatment-control difference when there is no true program effect (known as Type I error). Therefore, in assessing whether a statistically significant treatment-control difference, especially one with a p-value between .05 and .10, should be interpreted as a true program impact, we consider whether the sign and magnitude of the predicted effect are consistent with those for related outcomes.

Despite our large sample of patient episodes, it is unlikely that we will be able to identify small treatment-control differences as statistically significant, because design effects greatly reduce the precision of our estimates. For example, ignoring the design effects associated with our weighted sample and the clustering of episodes within agencies, the minimum detectable effect of the demonstration on the number of visits is about 2 percent under a two-tailed test at the 10 percent significance level, with 80 percent power. After accounting for design effects, however, the minimum detectable effects are about 11 percent. Thus, despite our large sample size, it is unlikely

---

[13]Two-tailed tests are used throughout this analysis. Although we expect that program impacts will be to reduce the use of most home health services, that is not universally true, since agencies could increase some types of visits while reducing others to hold down costs. Thus, to avoid confusion over whether a given test is two-tailed or one-tailed, and to flag estimates of the "wrong" sign that are large enough to be statistically significant, all t-tests conducted are two-tailed.

that we would detect the effects of prospective payment on overall visits, or outcomes, unless they were at least moderate in size.

4. Weighting

As noted previously, we weight the episodes in the main regression analysis to give agencies equal representation in the analysis. We use this approach for two reasons. First, because the demonstration is implemented at the agency level (not the episode level), the agency is the behavioral unit of interest. Second, the use of weighted data ensures that the impact estimates will not be dominated by the experiences of a few large agencies.

For each agency i, we construct the "agency equal" weight as follows:

$$(4) \quad w_i = \frac{1/n_i}{k/n},$$

where n_i is the number of (episode-level) observations in agency i, k is the number of agencies, and n is the total number of observations for all agencies. The weights range from a high of 24.8 to a low of 0.14, with 75 percent of the agencies having weights between 7.1 and 0.5.

While we prefer this weighting approach, one potential drawback is that the overall impact estimates may be distorted by placing very large weight on episodes in small agencies, which may include outliers. Therefore, in the sensitivity analysis, we also examine the impacts of the demonstration when giving each agency representation in the analysis that is equal to its relative size. Since all the agency's episodes during the demonstration are included in the analysis, this normally would be equivalent to conducting the analysis without sample weights; however, for this interim report, our sample includes home health admissions occurring over a different length of time

for each agency.[14] Thus, in order to reflect the agencies' relative size accurately, we must scale each observation for an agency by the (relative) time that it had been in the demonstration by the time the file was cut off (August 1996).

For each agency i, this "agency share" weight is given by

$$(5) \quad w_i{}^s = \frac{\overline{t}}{t_i},$$

where t_i is the length of time that agency i has been in the demonstration as of August 1996, and $\overline{t}$ is the average of the t_i's across agencies (approximately 10 months).

5. Design Effects

To draw appropriate inferences about the expected effects of a national prospective payment program, our estimated standard errors must reflect the fact that our episode observations are clustered in a small number of agencies. The variances of the impact estimates generated from standard statistical packages account for the number of episodes included in the sample, treating them as though they are a random sample from an infinite population of episodes.[15] They do not, however, account for the fact our sample consists of episodes from a limited number of agencies.

To correct for possible nonindependence of observations within agencies, we use SUDAAN software to obtain the appropriate standard errors for our impact estimates. The SUDAAN

---

[14]We currently have data over varying time intervals because agencies entered the demonstration at the start of their fiscal year, which differs by agency. Only episodes beginning on or before August 31, 1996, are included in the analysis. For the final report, we will have up to two years of data on all agencies.

[15]Our sample actually includes the population of patient episodes taking place in demonstration agencies over the early demonstration period. However, because we wish to make inferences about the outcomes for patients admitted in other times and to other agencies, we treat episodes in the data as though they were drawn in a simple random sample from the pool of all (future) episodes in all agencies.

calculations also account for the greater variance introduced by using sample weights in the regression models.

6.   Robustness Checks

While we expect the regression models used in the main analysis to be robust, they may be sensitive to three important factors. First, as noted previously, the use of sample weights that equate agencies' representation in the data may give small agencies undue influence in the analysis. Second, the general existence of outliers, either within or across agencies, may have undue influence on the analysis.   Third, the censoring of continuous variables at a discrete value may introduce statistical bias into the OLS models.

Alternative Sample Weighting. To investigate the sensitivity of our results to the weighting approach, we examine the impacts of prospective payment on key outcomes using the "share of episode" weights described previously. To the extent that the impact estimates are similar under the two approaches, it strongly suggests that our results are not overly influenced by a small number of anomalous observations from agencies providing few episodes.   Thus, the results may be more broadly interpreted for policy purposes. Conversely, while dissimilar results under the two sample weights do not necessarily indicate that the main results are "incorrect," they do suggest that further assessment is required to determine the most valid influences for policy.

Outliers. While the distribution of our outcome variables is generally not highly skewed, it is still important to examine the degree to which our findings on service use are affected by a small number of outlier episodes.   We use two methods to investigate the effects of outliers. First, to account for outliers at the agency level, we estimate models excluding all episodes from eight agencies--with the two highest and two lowest mean values for a given outcome in both the control group and the treatment group. Since we use the sample weights that give each agency equal weight,

the removal of these eight outlier agencies leaves about 90 percent (77/85) of the weighted sample intact.

The second approach that we use is to redefine the dependent variable as the natural logarithm of the outcome measure and regress the log variable against the standard set of explanatory variables shown in equation (1). While this model effectively reduces the influence given to outliers in the simple linear model, it has two major limitations. First, it may also substantially understate the true demonstration effects if they take place at the upper end of the distribution. Second, the natural logarithm is undefined for values of zero (or less). As a result, we use this model only to examine the effect of outlier episodes on the results for total visits.

The log-linear model is given by:

$$(6) \quad lnY = \alpha + XB + \delta T + \epsilon.$$

We estimate the impact of prospective payment on the outcome Y through a transformation of the coefficient δ on the treatment status variable T:

$$(7) \quad \% \ change \ in \ Y = exp^\delta - 1.$$

Censoring of the Dependent Variable. The Tobit model will be used to obtain consistent estimates of impacts on outcomes where the dependent variable has been censored at a particular value.[16] By comparing the results from these models with those obtained from OLS regressions, we determine the potential sensitivity of our main findings to the effects of censoring.

---

[16] The consistency of the Tobit model rests on very strong statistical assumptions, and violation of these assumptions may lead to substantial bias in the estimated impacts. Therefore, we prefer to use this model only as a sensitivity test and rely on more robust OLS models for the main findings.

Variables may be censored from the left, right, or both sides. All variables for the number of visits of a particular type are censored on the left at a value of zero. The length of an episode (in this report) is censored on the right at 120 days, the upper bound of the at-risk period.

The Tobit model for our left-censored dependent variables (visits) assumes that an unobserved, underlying index of the need for services exists. If this index (Y^*) exceeds some threshold (zero in our specification), then the patient receives the needed amount of the services; if the index is less than the threshold, the patient receives no care. The model therefore consists of a probabilistic component for the likelihood of needing a particular type of service, and a linear component for the expected number of visits, conditional on having a need for services above the threshold:

$$(8) \quad Y^* = \alpha + XB + \delta T + \epsilon$$
$$Y = Y^* \; if \; Y^* > 0$$
$$Y = 0 \; if \; Y^* \leq 0$$

where the variables and parameters are the same as in equation (1).

After the parameters of the Tobit model have been estimated, they are used to obtain predictions of the expected value of the given outcome for each observation. This is given by the product of probability that during the episode a patient uses a given service and the amount of the service used, conditional on it being greater than zero. These predicted values are calculated twice: once assigning a given observation to the control group, and once assigning it to the treatment group. The mean difference between these two predicted values is our estimate of the impact of the demonstration on a given outcome.

The Tobit model to correct for right censoring of the episode length follows the same approach. There is continuous portion of the episode distribution from 1 to 119 days, and a discrete mass point at 120 days. The model is therefore given by

$$
\begin{aligned}
(9) \quad Y^* &= \alpha + X\mathrm{B} + \delta T + \epsilon \\
Y &= Y^* \; \text{if } Y^* < 120 \\
Y &= 120 \; \text{if } Y^* \geq 120.
\end{aligned}
$$

After the parameters of the Tobit model have been estimated, they are used to obtain predictions as before. The predicted length within the first 120 days for an episode is given by the sum of two items: (1) the product of the predicted probability that the episode lasts less than 120 days and the predicted value of episode length, conditional on it being less than 120 days; and (2) the product of the probability that an episode is 120 days or longer and 120, the maximum value possible. These predicted values are estimated twice for each observation, once with the case assigned to the control group and once without it assigned to the treatment group. The mean difference between these two predicted values is our estimate of the impact of the demonstration on the episode length during the "at-risk" period.[17]

---

[17]This analysis of episode length conducted here is only for exploratory purposes, since actual episode lengths longer than 120 days can be observed in our data. Future analyses of episode length will use the actual episode length and will have variable censoring points for those observations for which the end of the episode is not observed during the follow-up period available.

III. IMPACTS ON THE USE OF SERVICES IN THE FIRST 120 DAYS

A. EXPECTED EFFECTS OF PROSPECTIVE PAYMENT

The major method by which agencies are expected to generate profits in the per-episode payment demonstration is reducing the number of visits they provide during the 120-day at-risk period. This could be accomplished by reducing visits to the same mix of patients or by seeking a mix of patients needing less care.

Agencies have two ways of reducing the average number of visits per episode relative to what they would have provided to the same patients under traditional-cost-based reimbursement: (1) the interval between visits (that is, providing fewer visits per week) without reducing the length of the episode, and (2) eliminating visits near the end of care and discharging the patient earlier. In either case, treatment agencies might seek care from patients' relatives, home- and community-based services, or another home health agency to substitute for visits they were no longer providing.[1] In addition, for patients who are discharged earlier, hospice care or nursing home care might replace home health care. Treatment agencies might also eliminate visits in favor of telephone monitoring, which would replace visits during the episode or replace the last visits in an episode. Since claims data do not include information on telephone encounters, we could observe less-frequent visits and/or a reduction in episode length, depending on the timing of telephone monitoring.

Early in the demonstration, treatment agencies reported implementing a number of different types of activities to reduce visits (see Phillips and Thompson 1997). Some of these activities, such as greater supervision of care planning and greater use of care maps, might lead either to a reduction

---

[1] Under Medicare regulations, two home health agencies are allowed to provide services simultaneously.

in the frequency of visits (without affecting episode length) or to earlier discharge. Other activities seem likely to lead to reduced episode length. Examples include the use of specialists to speed healing of wounds and more-frequent visits early in the episode to speed learning of self-care routines.

A treatment agency could also reduce its average number of visits during the 120-day period by seeking a mix of patients needing less care than the mix of patients the agency had served in the past. However, in the site visits, we found no evidence that treatment agencies did so. Rather, agencies reported that, to maintain a flow of new business in an extremely competitive environment, they must be highly responsive to the needs of their referral sources. They are not in a position to refuse patients even if these patients are expected to require extensive care. Moreover, it seems likely that agencies would have to alter either their referral sources or their intake procedures in order to shift to a less costly mix of patients. In the site visits, we found no evidence that agencies changed their referral sources or intake procedures in the early months of the demonstration in response to per-episode payment (Phillips and Thompson 1997). Finally, even if a demonstration agency were able to attract a mix of patients with less need for care than the ones previously served, it would not necessarily make larger profits. The case-mix adjuster would reduce the per-episode payments to the agency if the new cases were more concentrated than the predemonstration cases in the lower-cost rate cells. This provides a further drawback to seeking profits by altering case mix.

B. DID PROSPECTIVE PAYMENT AFFECT THE NUMBER OF VISITS RENDERED?

The demonstration payment method gives agencies a clear incentive to provide fewer visits during the first 120 days of care. If prospectively paid (treatment) agencies can reduce the number of visits without proportionately increasing unit costs, they will make a profit. Agencies reimbursed on the conventional cost basis (controls) have no financial incentive to reduce visits.

The impacts of prospective payment may be somewhat sensitive to the level of managed care provided by demonstration agencies. For agencies with a large managed care caseload, efforts to reduce services for managed care patients may spill over to affect patients paid on a fee-for-service basis. While these spillover effects are likely to be weak, to the extent they exist, our impact estimates may be smaller the more managed care patients that demonstration agencies serve. Based on our site visit data, the demonstration agencies served relatively few managed care patients; thus, we expect that our impact estimates will be largely unaffected by the influence of managed care.

We do, in fact, find large and significant treatment-control differences in the number of visits provided during the first 120 days of care (Table III.1). Overall, prospective payment led treatment agencies to provide an average of 7.9 fewer visits than control agencies. Relative to the mean of 45 visits per episode for the control group, this reflects a 17 percent decline in total visits.

This overall decline was primarily the result of fewer visits by both skilled nurses and home health aides. These two visit types together accounted for 37 (83 percent) of the 45 total visits per episode in the control agencies. Prospectively paid agencies provided an average of four fewer skilled nursing visits and about three fewer home health aide visits per episode during the at-risk period, both statistically significant declines of about 18 percent. Prospective payment also led to a statistically significant drop of 0.3 visits (per at-risk period) by medical social workers, a 37 percent decline relative to the mean for control group agencies. Occupational and physical therapy also showed fairly large proportional declines, although they were smaller than the effects for skilled nursing and home health aides and were not statistically significant. The weaker results for therapy visits are consistent with expectations that such services are less likely than aide and nursing visits

TABLE III.1

THE IMPACT OF PER-EPISODE PAYMENT ON THE NUMBER
OF VISITS IN FIRST 120 DAYS
(Agencies Weighted Equally)

	Control Group Mean	Impact[a] (p-value)	Impact as a Percentage of the Control Group Mean
Total Visits	45.03	−7.85 (.00)	−17.4
Skilled Nursing Visits	21.71	−3.97 (.00)	−18.3
Home Health Aide Visits	15.60	−2.86 (.01)	−18.3
Physical Therapy Visits	5.63	−0.55 (.29)	−9.8
Occupational Therapy Visits	1.01	−0.25 (.12)	−24.8
Speech Therapy Visits	0.30	0.07 (.25)	23.3
Medical Social Worker Visits	0.78	−0.29 (.01)	−37.2

SOURCE: Medicare Claims Data.

Sample Size: 25,561 episodes in 44 treatment agencies; 25,130 episodes in 41 control agencies.
                Episodes are weighted so each agency has equal weight.

[a]These estimates, obtained from regression models, represent treatment-control differences in the
expected number of visits per episode. The p-values are based on estimated standard errors that
account for the effects of clustering and weighting.

to be oversupplied under cost-based reimbursement due to shortages of personnel and the fact that many agencies obtain therapy services by contracting with other organizations.

Decomposition of these impacts indicates that prospective payment had little effect on the likelihood of receiving a specific type of visit during the at-risk period; however, conditional on receiving a visit, it led to large reductions in the number of visits rendered (Table III.2).[2] Relative to the control group mean, the only large and statistically significant impact on the probability of a visit was for occupational therapy, which fell nearly 35 percent. The probability of a skilled nursing visit also fell, but the decline was quite small. In contrast, for episodes that received one or more visits of a given type, prospective payment led to large and significant declines in visits (relative to cost-based reimbursement) by skilled nurses, home health aides, physical therapists, and medical social workers. The most dramatic of these declines was in home health aide visits, which were 5.9 visits lower on average under prospective payment than under cost-based reimbursement--a decline of about 17 percent relative to the control group mean.

The impacts on skilled nursing, physical therapy, and home health aide visits are largely consistent with expectations. During the site visits, treatment agencies described how they were changing their approach toward treating patients from "taking care of patients" to "making the patients as well and as independent as soon as possible." In particular, agencies identified a number of initiatives they were using that should decrease service use, including greater supervision of care planning, the application of care maps or critical pathways, heavier reliance on family and community supports, and increased use of telephone contacts in place of visits. Agencies used these

_____

[2]This decomposition refers to the two-part model specification introduced by Duan et al. (1983). The model accounts for the censoring of visits at zero by distinguishing the effect of prospective payment on the probability of receiving any visits of a given type from its impact on the number of visits, conditional on having received at least one.

TABLE III.2

DECOMPOSITION OF THE IMPACT OF PER-EPISODE PAYMENT
ON THE NUMBER OF VISITS IN FIRST 120 DAYS
(Agencies Weighted Equally)

Service Type	Probability of Receiving the Service (Percentage)		Number of Visits (If Service Is Received)	
	Control Group Mean	Impact[a] (p-value)	Control Group Mean	Impact[b] (p-value)
Skilled Nursing Visits	96.5	−2.6 (.03)	22.5	−3.8 (.00)
Home Health Aide Visits	45.5	−0.3 (.88)	34.3	−5.9 (.00)
Physical Therapy Visits	41.8	−0.5 (.84)	13.5	−1.5 (.03)
Occupational Therapy Visits	12.6	−4.4 (.01)	8.0	−0.3 (.59)
Speech Therapy Visits	3.2	−0.1 (.81)	9.4	1.1 (.17)
Medical Social Worker Visits	28.6	−2.8 (.34)	2.7	−0.8 (.00)

SOURCE: Medicare Claims Data.

Sample Size: 25,561 episodes in 44 treatment agencies; 25,130 episodes in 41 control agencies.
              Episodes are weighted so each agency has equal weight.

[a]These estimates, obtained from logistic regression models, represent treatment-control differences
in the probability that an episode includes at least one visit. The p-values are based on standard
errors adjusted to account for the effects of clustering and weighting.

[b]These estimates, obtained from ordinary least squares regression models, represent treatment-control
differences in the average number of visits, for episodes with at least one visit of the type in
question. The p-values take into account the effects of clustering and weighting on the underlying
standard errors.

techniques to reduce all types of services but focused on decreasing skilled nursing and home health aide visits--precisely where we observe the largest reduction in visits. However, some of these techniques are, in practice, particularly relevant to the reduction of nursing and aide visits (for example, increased telephone contact and heavier reliance on family and community supports).

The large proportionate decline in the use of medical social workers, however, conflicts with information gathered during the site visits. Some treatment agencies indicated that they increased the use of medical social workers to arrange for needed community support services and to help patients who need nursing home placement decide more quickly to seek it. The overall decrease in medical social worker visits by treatment group agencies suggests that any such increases were outweighed by the general incentive to reduce the cost of care.

C. DID PROSPECTIVE PAYMENT AFFECT THE EPISODE COMPOSITION?

In addition to reductions in the overall number of visits per at-risk period, agencies may be able to cut costs substantially by changing the mix of services provided to patients. Such a change could occur in several ways. Agencies may, for example, attempt to substitute away from high-cost visits by skilled nurses toward relatively low-cost visits by home health aides. Or they may be more likely to use medical social workers earlier in an episode to hasten a patient's exit from home health into a more permanent, long-term care environment or into community-based care by informal caregivers. Alternatively, the mix of visits may be affected if agencies are able to reduce certain types of visits by a larger proportion than others without adversely affecting patient outcomes. Thus, we investigate whether prospective payment changed the composition of a home health episode.

Prospective payment had no discernible effect on the mix of home care services (Table III.3). The proportion of total visits provided by each of the six types of home health staff differed little for

TABLE III.3

THE IMPACT OF PER-EPISODE PAYMENT ON THE TYPES
OF CARE RENDERED IN FIRST 120 DAYS
(Agencies Weighted Equally)

	Control Group Mean	Impact[a] (p-value)
Percentage of Visits by Skilled Nurses	60.6	−0.4 (.82)
Percentage of Visits by Home Health Aides	20.0	−0.1 (.91)
Percentage of Visits by Therapists	17.1	1.0 (.52)
Percentage of Visits by Medical Social Workers	2.2	−0.5 (.10)
Number of Days Until the First Medical Social Worker Visit[b]	20.8	−2.8 (.02)

SOURCE: Medicare Claims Data.

Sample Size: 25,561 episodes in 44 treatment agencies; 25,130 episodes in 41 control agencies.
Episodes are weighted so each agency has equal weight.

[a]These estimates, obtained from ordinary least squares regression models, represent treatment-control (percentage point) differences for each of the outcomes listed in the first column. The p-values are based on standard errors adjusted to account for the effects of clustering and weighting.

[b]The sample size for this outcome is 12,960. Bill records do not identify the date of service for a social worker visit. This date has been approximated by using the mid-point of the billing period on which the first medical social worker visit was recorded.

treatment and control groups, and none of these differences was statistically significant from zero.[3] This is consistent with the information collected in the site visits, where few agencies even considered substituting one type of provider for another. However, the finding of no effects on the proportionate mix of visits is somewhat surprising. Our expectation was that agencies would reduce the proportion of aide visits, and perhaps of skilled nursing visits, more than that of therapies. The point estimates are consistent with this expectation--reductions in the average number of physical therapy visits (11 percent) and occupational therapy visits (4 percent) were smaller than the 18 percent reductions for both nursing and aide visits. Nonetheless, the differences were small enough that they did not significantly affect the proportions of total visits per episode accounted for by the different visit types.

Information from our site visits did suggest one important change in the provision of care: agencies were including social workers visits much earlier in the episode. Treatment agencies often indicated that under the prospective payment, they focused more on planning for discharge from the start of the episode; as part of this effort, they sent the medical social worker into the patient's home as early as possible to smooth the discharge process. This allowed the social worker to put post Medicare services in the home and make sure there were no delays in discharge. In contrast, the control agencies rarely indicated that they focused on discharge from the outset of the episode, except perhaps for their managed care patients.

---

[3]Each proportion variable that we examine reflects the mean of the individual (episode-level) proportion of visits provided of a given type; this is not equivalent to the overall proportion of visits provided of a given type (which can be calculated from Table III.1). For example, the mean proportion of skilled nursing visits provided to individual patients was 60.6 percent; however, the percentage of skilled nursing visits provided overall in our data was 48 percent.

Our impact estimates strongly support the site visit findings. For those patients in the control group who received a medical social worker visit, the average length until the first visit was about three weeks. In response to prospective payment, treatment agencies reduced this time until the first medical social worker visit by nearly three days, about 13 percent of the control group mean.[4]

D. DID PROSPECTIVE PAYMENT AFFECT THE EPISODE LENGTH?

If agencies have reduced the volume of care provided as a result of the payment method, we also might expect the demonstration to reduce the average length of time over which care is delivered during the 120-day period. Alternatively, if agencies reduce visits per episode mainly by reducing the frequency of visits, there may be little or no effect on episode length. One advantage of discharging the patient more quickly is that agencies can reduce somewhat the greater overhead costs associated with longer episodes of care, such as bimonthly reassessments, and earn a greater profit. Agencies might decrease the observed length of an episode simply by ending visits sooner and compensating with some further action, such as teaching patients self-care techniques that would let then fend for themselves, placing them in an alternative health setting (a nursing home, for example), training informal caregivers for them, or substituting telephone calls for the last few visits. Agencies may also continue to serve the patient, but not with Medicare funding, by finding an alternative payment source (such as Medicaid).

---

[4]Unfortunately, bill records do not identify the date of service for social worker visits, leading us to estimate the actual date of service as the mid-point of the billing period in which the first visit was recorded. It is therefore possible that our impact estimate may be a product of measurement error in the date of service. This possibility is remote, however, since there is no reason to suspect that the "average" measurement error would differ between treatment and control agencies. Moreover, our impact estimate is consistent with information from site visits and with demonstration incentives.

Consistent with its impact on the volume of care provided, prospective payment dramatically reduced the length of an episode (Table III.4). Relative to the average episode length within the at-risk period for control agencies (about 70 days), the episode length for treatment agencies was 10 days shorter because of prospective payment--a statistically significant decline of 14 percent.[5] Consistent with this impact, the probability that an episode would end during the first month (1 to 31 days) showed a dramatic increase of 8.1 percentage points as a result of prospective payment. Despite this increase, however, there was essentially no change in the probability of an episode ending during the second, third, or, fourth month of the at-risk period.[6] Instead, the entire increase in the probability of episodes ending within the first month was offset by a sharp decline in the rate of episodes ending beyond the at-risk period (after 120 days). Prospective payment reduced the percentage of episodes ending after 120 days by nearly 10.7 percentage points--a decline of nearly one-third relative to the control group mean of 33.9 percent.

The formation of this treatment-control difference in the rate of completed episodes may be seen clearly from Figure III.1. As the episode length increases up until 31 days, treatment agencies have episodes ending at a steadily higher rate than controls. This leads to the gap between treatment and controls of about 8 percentage points by 31 days. After 31 days, the gap continues to widen to about 10 percentage points by 48 days, and it then remains roughly constant until the end of the episode payment at 120 days.

---

[5] The maximum episode length within the at-risk period is 120 days.

[6] The treatment-control differences shown in the probability of an episode ending during a given time period are based on a comparison non-(regression)-adjusted means. Estimates from a multinomial logit model (not shown) that controls for agency and patient characteristics were highly consistent with these unadjusted impacts. We prefer to present unadjusted results because of the difficulty of recovering correct standard errors on the impacts under the multinomial logit model.

TABLE III.4

THE IMPACT OF PER-EPISODE PAYMENT ON EPISODE LENGTH
(Agencies Weighted Equally)

	Control Group Mean	Impact[a] (p-value)
Length of an Episode Within the At-Risk Period (Days)[b]	69.9	-9.8 (.00)
Probability of an Episode Ending:		
During the First Month (1 to 31 Days) (Percentage)	27.5	8.1 (.00)
During the Second Month (36 to 62 Days) (Percentage)	24.3	1.1 (.45)
During the Third Month (63 to 92 Days) (Percentage)	7.5	0.7 (.30)
During the At-Risk Period of the Fourth Month (93 to 120 Days) (Percentage)	6.8	0.8 (.16)
After the At-Risk Period (More than 120 Days) (Percentage)	33.9	-10.7 (.00)

SOURCE: Medicare Claims Data.

Sample Size: 25,561 episodes in 44 treatment agencies; 25,130 episodes in 41 control agencies.
           Episodes are weighted so each agency has equal weight.

[a]These estimates represent the treatment-control difference in the mean for each outcome listed in the first column. The p-values are based on standard errors adjusted to account for the effects of clustering and weighting.

[b]Maximum value for episode length within the at-risk period is 120 days. The impact shown is derived from an ordinary least squares model.
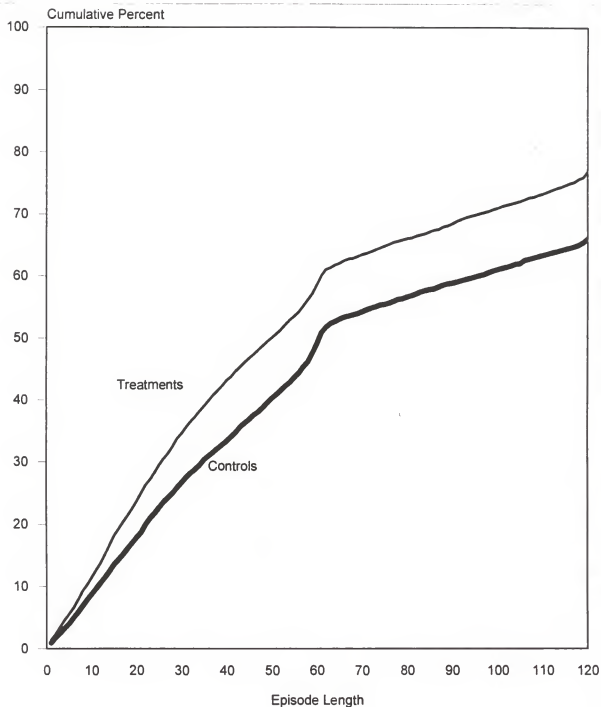
   *Significantly different from zero at the .10 level, two-tailed test.
  **Significantly different from zero at the .05 level, two-tailed test.
***Significantly different from zero at the .01 level, two-tailed test.

FIGURE III.1

EPISODE LENGTH BY TREATMENT STATUS
(AGENCIES WEIGHTED EQUALLY)

Source: Medicare Claims Data.

The results from Table III.4 and Figure III.1 suggest a very simple finding: the demonstration payment method consistently shortened episodes of all different lengths. That is, as prospective payment led episodes to end in the last month of the at-risk period rather than in the outlier period, it led other episodes to end in the third month rather than the fourth, the second month rather than the third, and so on. The net result, consistent with Figure III.1, is that treatment agencies had a far smaller percentage of long episodes (beyond 120 days), a "piling up" of relatively short episodes (less than one month), and little change in the rate of episodes of moderate length (between one month and 120 days).[7]

E. DID PROSPECTIVE PAYMENT AFFECT AGENCY SUBGROUPS DIFFERENTLY?

Agencies are likely to respond to the demonstration incentives differently depending on their characteristics. We focus on four characteristics that we believe are most likely to affect agency behavior — for-profit status, predemonstration practice (use) patterns, size, and auspice. For each of these characteristics, we form respective subgroup pairs as follows: (1) a for-profit subgroup and a non-profit subgroup; (2) a "high-use practice pattern" subgroup for agencies with a base year practice pattern index value above the median value, and a "low-use practice pattern" subgroup for agencies at or below the median value; (3) a "small size" subgroup for agencies with fewer than 30,000 visits in the base year, and a "large size" subgroup for agencies with more than 30,000 visits; and (4) a hospital-based subgroup and a freestanding subgroup. We then examine differences between subgroup pairs in each of the following outcomes: total number of visits, proportion of

---

[7]These results are also consistent with an alternative hypothesis that prospective payment simply reduced long episodes (beyond 120 days) to short episodes (within 31 days). This hypothesis is less plausible, however, and agencies never mentioned it during our interim visits.

skilled nursing visits, proportion of home health aide visits, and length of an episode within the at-risk period.

Proprietary agencies have the clearest incentive to decrease visits under the demonstration. Nonprofit agencies have more-diverse objectives; while they would like to earn a surplus to support their charity work and computerization, they may reduce visits by a smaller fraction than for-profits because of greater concern for patients' and informal caregivers' well-being. Prospective payment could result in no change in utilization for nonprofit agencies, especially in light of the loss protection available under the demonstration. As a result, the responses to the demonstration from for-profit and nonprofit agencies may differ substantially.

Because they have made less effort to control utilization in the past, agencies with preexisting high-use practice patterns may be able to respond more aggressively to the demonstration payment method. In particular, those agencies that have not faced pressure from managed care organizations (and other competitive market forces) have had little incentive to contain costs under cost-based reimbursement. Under prospective payment, these agencies may therefore be more capable of making large-scale reductions in the amount of care that they provide per episode.

We also expect that smaller agencies (those under 30,000 visits per year) may respond differently to the demonstration incentives, although the direction of that response is uncertain. On the one hand, smaller agencies may find it easier to communicate with their staff and implement changes in care patterns more quickly, bringing about larger impacts among the smaller agencies, at least during this initial period of the demonstration. On the other, small agencies may find it less financially rewarding to reduce visits, because it may be more difficult to reduce their fixed costs.

Hospital-based agencies have different financial incentives, so they may respond to prospective payment differently. These agencies have the incentive to maximize profits for the entire hospital

system, so their practices may differ from those of freestanding agencies. For example, a hospital-based system fearful that physicians will admit their patients to other hospitals if the home health agency noticeably reduces visits may not respond as strongly to the demonstration incentives.

Surprisingly, we find very little difference in the response of various types of agencies to prospective payment (Tables III.5 to III.8). Program impacts did not vary consistently with any of the four agency characteristics used to define subgroups, suggesting that agencies of different types responded similarly to the demonstration incentives. In almost all cases, the results for the subgroups mimic the overall results: the treatment group agencies reduced the number of visits and lengths of episodes but did not change the mix of visit types provided during episodes. The fact that the magnitudes of the impacts estimates for each subgroup are roughly the same is consistent with the nonsignificant interaction effect.

One exception to the generally consistent findings across agency subgroups is the effect of prospective payment on the proportion of skilled nursing visits provided by nonprofit and for-profit agencies. Not only do we find a strongly significant difference in the impacts between these two subgroups, but we also find that the impacts are in opposite directions. While for-profit agencies significantly reduced the proportion of skilled nursing visits per episode, nonprofit agencies significantly increased this proportion. This suggest that the overall finding of no change in the proportion of skilled nursing visits per episode (shown in Table III.3) is actually due to offsetting effects among proprietary and nonprofit agencies, which are equally represented in our sample.

One may have expected to observe a decline in the proportion of visits that are for skilled nursing, because treatment agencies have an incentive to substitute away from skilled nurses in favor of low-cost home health aides. Such substitution, however, may be dominated by several other factors. In particular, since a much smaller number of therapy and medical social worker visits are

TABLE III.5

IMPACT OF PER-EPISODE PAYMENT ON THE USE OF SERVICES,
BY WHETHER THE AGENCY IS FOR-PROFIT OR NONPROFIT
(Agencies Weighted Equally)

	Control Group Mean	Impact[a] (p-value)[b]
Total Visits		
For-profit	51.6	−8.0***
Nonprofit	38.2	−7.3***
p-value for impact difference between subgroups		(.87)
Percentage of Visits by Skilled Nurses		
For-profit	62.2	−4.6**
Nonprofit	59.0	3.7*
p-value for impact difference between subgroups		(.01)
Percentage of Visits by Home Health Aides		
For-profit	21.9	−1.4
Nonprofit	18.0	1.3
p-value for impact difference between subgroups		(.27)
Length of Episode Within the At-Risk Period (Days)		
For-profit	73.8	−10.2***
Nonprofit	65.8	−9.9***
p-value for impact difference between subgroups		(.96)

SOURCE: Medicare Claims Data.

		For-Profit		Nonprofit	
		Episode	(Agencies)	Episode	(Agencies)
Sample Size:	Treatment	5,181	(21)	20,380	(23)
	Control	6,401	(21)	18,729	(20)

[a]These estimates, obtained from ordinary least squares regression models, represent treatment-control differences for each of the outcomes listed in the first column.

[b]All p-values are based on standard errors adjusted to account for the effects of clustering and weighting. The p-value shown for each outcome corresponds to a test of whether the impacts for the two subgroups are statistically different from one another. Asterisk(s) shown next to a given impact indicate that it is statistically significant from zero.

*p-value < .10, two-tailed test.
**p-value < .05, two-tailed test.
***p-value < .01, two-tailed test.

TABLE III.6

IMPACT OF PER-EPISODE PAYMENT ON THE USE OF SERVICES,
BY WHETHER THE AGENCY HAD A HIGH-USE
OR LOW-USE PRIOR PRACTICE PATTERN
(Agencies Weighted Equally)

	Control Group Mean	Impact[a] (p-value)[b]
Total Visits		
High-use practice pattern	53.1	−10.7***
Low-use practice pattern	32.5	−4.6**
p-value for impact difference between subgroups		(.04)
Percentage of Visits by Skilled Nurses		
High-use practice pattern	60.2	−1.0
Low-use practice pattern	61.3	0.3
p-value for impact difference between subgroups		(.71)
Percentage of Visits by Home Health Aides		
High-use practice pattern	23.5	−0.7
Low-use practice pattern	14.6	0.6
p-value for impact difference between subgroups		(.58)
Length of Episode Within the At-Risk Period (Days)		
High-use practice pattern	74.3	−10.7***
Low-use practice pattern	63.1	−9.4***
p-value for impact difference between subgroups		(.74)

SOURCE: Medicare Claims Data.

		High-Use		Low-Use	
		Episode	(Agencies)	Episode	(Agencies)
Sample Size:	Treatment	12,117	(18)	13,444	(26)
	Control	9,788	(25)	15,342	(16)

NOTE: An agency is defined as having a high-use practice pattern if its (case-mix adjusted) number of visits per episode in the base (predemonstration) period was above the median for all agencies in the sample. Otherwise, it is defined as having a low-use practice pattern.

[a]These estimates, obtained from ordinary least squares regression models, represent treatment-control differences for each of the outcomes listed in the first column.

[b]All p-values are based on standard errors adjusted to account for the effects of clustering and weighting. The p-value shown for each outcome corresponds to a test of whether the impacts for the two subgroups are statistically different from one another. Asterisk(s) shown next to a given impact indicate that it is statistically significant from zero.

 *p-value < .10, two-tailed test.
 **p-value < .05, two-tailed test.
***p-value < .01, two-tailed test.

TABLE III.7

IMPACT OF PER-EPISODE PAYMENT ON THE USE OF SERVICES,
BY WHETHER THE AGENCY IS SMALL OR LARGE SIZE
(Agencies Weighted Equally)

	Control Group Mean	Impact[a] (p-value)[b]
Total Visits		
Small size	58.7	−10.5**
Large size	41.7	−6.7***
p-value for impact difference between subgroups		(.44)
Percentage of Visits by Skilled Nurses		
Small size	59.4	2.5
Large size	60.9	−1.4
p-value for impact difference between subgroups		(.30)
Percentage of Visits by Home Health Aides		
Small size	27.2	0.1
Large size	18.3	−0.1
p-value for impact difference between subgroups		(.93)
Length of Episode Within the At-Risk Period (Days)		
Small size	81.6	−14.4***
Large size	67.1	−8.4***
p-value for impact difference between subgroups		(.21)

SOURCE: Medicare Claims Data.

		Small Size		Large Size	
		Episode	(Agencies)	Episode	(Agencies)
Sample Size:	Treatment	1,769	(15)	23,791	(29)
	Control	658	(8)	24,472	(33)

NOTE: Large (small) agencies are defined as those that provided more (fewer) than 30,000 visits in the predemonstration year.

[a]These estimates, obtained from ordinary least squares regression models, represent treatment-control differences for each of the outcomes listed in the first column.

[b]All p-values are based on standard errors adjusted to account for the effects of clustering and weighting. The p-value shown for each outcome corresponds to a test of whether the impacts for the two subgroups are statistically different from one another. Asterisk(s) shown next to a given impact indicate that it is statistically significant from zero.

 *p-value < .10, two-tailed test.
 **p-value < .05, two-tailed test.
***p-value < .01, two-tailed test.

TABLE III.8

IMPACT OF PER-EPISODE PAYMENT ON THE USE OF SERVICES,
BY WHETHER THE AGENCY IS HOSPITAL-BASED
OR FREESTANDING
(Agencies Weighted Equally)

	Control Group Mean	Impact[a] (p-value)[b]
Total Visits		
Hospital-based	34.3	−8.6***
Freestanding	46.9	−7.6**
p-value for impact difference between subgroups		(.72)
Percentage of Visits by Skilled Nurses		
Hospital-based	62.0	−4.0
Freestanding	60.4	0.1
p-value for impact difference between subgroups		(.38)
Percentage of Visits by Home Health Aides		
Hospital-based	16.4	0.8
Freestanding	20.6	−0.2
p-value for impact difference between subgroups		(.74)
Length of Episode Within the At-Risk Period (Days)		
Hospital-based	61.5	−14.5**
Freestanding	71.4	−9.5***
p-value for impact difference between subgroups		(.43)

SOURCE: Medicare Claims Data.

		Hospital Based		Freestanding	
		Episode	(Agencies)	Episode	(Agencies)
Sample Size:	Treatment	2,139	(4)	23,422	(40)
	Control	3,870	(6)	21,260	(35)

[a]These estimates, obtained from ordinary least squares regression models, represent treatment-control differences for each of the outcomes listed in the first column.

[b]All p-values are based on standard errors adjusted to account for the effects of clustering and weighting. The p-value for each outcome corresponds to a test of whether the impacts for the two subgroups are statistically different from one another. Asterisk(s) shown next to a given impact indicate that it is statistically significant from zero.

*p-value < .10, two-tailed test.
**p-value < .05, two-tailed test.
***p-value < .01, two-tailed test.

provided per-episode. any reduction in these services may increase the proportion of skilled nursing visits even if substitution were taking place.[8]

Among for-profit agencies, the decline in the proportion of skilled nursing visits suggests that substitution may have taken place; however, the small (insignificant) decline in home health aide visits contradicts this explanation. Instead, it seems that for-profit agencies were simply far more able to make reductions in visits by skilled nurses and home health aides than other services. This result is perhaps not surprising. Since for-profit agencies provided a large number of visits by skilled nurses and aides, we might expect these services to be far easier to reduce.[9] For the nonprofit agencies, the increase in the relative use of skilled nurses again suggests that little or no substitution took place. Moreover, it indicates that because nonprofit agencies provided relatively fewer skilled nursing and aide visits under fee-for-service, the nonprofits were unable to make disproportionately large cuts in these services. It is well known that for-profit agencies supply substantially more visits per episode than nonprofit agencies, and most of these additional visits are likely to be nursing and aide visits.[10] Thus, these are the services that for-profit agencies would most likely cut back on.

---

[8]Consider the following example. In the absence of the demonstration, agency X would have had an average episode with 50 visits, where skilled nurses provided 40 of these visits and therapists provided the remaining 10. Suppose that as a result of the demonstration, agency X cut 5 skilled nursing visits and 4 therapy visits, and in addition, it substituted 2 skilled nursing visits with 2 home health aides visits. Despite larger cuts in skilled nursing, the proportion of skilled nursing visits for agency X would actually increase from 80 percent (40/50) per episode to 83 percent (33/40).

[9]Among control agencies, for-profits provided about 40 percent more skilled nursing and aide visits than nonprofits. Moreover, as shown in Table III.5, an average of 84 percent of the visits per episode by for-profit control agencies were done by skilled nurses and home health aides, compared to 77 percent among nonprofit control agencies.

[10]In our data, for-profit agencies provided an average of about 40 percent more skilled nursing visits and home health aide visits than nonprofit agencies, far greater than any differences for other visit types.

As Table III.6 shows, a second, and perhaps more important, difference in impacts takes place between agencies defined by their predemonstration practice patterns. As expected, treatment group agencies that previously provided a larger-than-average number of visits (controlling for case mix) reduced the number of visits per episode during the demonstration by a significantly larger number (and percentage) of visits than agencies with a history of low use. High-use treatment group agencies reduced visits per episode by nearly 11 visits (about 20 percent), whereas low-use treatment agencies reduced them by fewer than 5 visits (about 15 percent). Among control group agencies, the high-use agencies delivered about two-thirds more visits per episode on average than the low-use group. This pattern suggests that high-use treatment group agencies may be able to reduce visits more than low-use ones in response to prospective payment, without adversely affecting patient outcomes. We find no differences between these two subgroups in impacts on the mix of visits provided or the length of the episode.

The observed difference in impacts on visits per episode for high- and low-use agencies masks large and significant differences in impacts on this outcome between for-profit and nonprofit agencies. That is, when impact estimates are allowed to vary only with whether an agency was for-profit or nonprofit (not shown), we find that reductions in visits per episode are much larger for the for-profit agencies. However, for-profit agencies tended to have much higher average visits per episode than nonprofits during the predemonstration period. Thus, when impact estimates are allowed to vary simultaneously with practice pattern and for-profit status, we find that it is the high-use/low-use practice pattern classification that explains differences in impacts across agencies. It would be incorrect, however, to conclude that the for-profit status of an agency has no bearing on its expected reduction in visits per episode under prospective payment, because such agencies are much more likely to have had high-use profiles under traditional cost-based reimbursement. One

may conclude, however, that if switched to prospective payment, for-profit agencies are no more likely than nonprofit agencies with similar practice patterns under cost-based reimbursement to reduce visits per episode.

F.   ROBUSTNESS OF ESTIMATED IMPACTS

To ensure that our estimates of program impacts on treatment-control differences are not heavily dependent on particular observations or statistical procedures, we perform several sensitivity tests.[11] First, we examine the dependence of our impact estimates on the regression model used by comparing our regression-based results to those obtained from the treatment-control difference in simple (weighted) means. Second, we investigate the effects of representing agencies in proportion to their share of episodes provided (rather than weighting all agencies equally). Third, we examine the sensitivity of our results to outliers: extreme values of visits per episode for particular patients and agencies. Fourth, we discuss how the estimated impacts differ with an analytic model (Tobit) that accounts for censoring of the dependent variable. Finally, we examine whether our results may be affected by unobserved factors not controlled for through the regression models.

1.   Comparison of Regression-Adjusted and Unadjusted Demonstration Impacts

As a first step to determine the dependence of our impact estimates on the regression model used, we compare the treatment-control difference in simple (weighted) mean visits per episode (Table III.9) to the estimates obtained from the regression (Table III.1). The difference in raw means (-11 visits) is substantially (40 percent) greater (in absolute value) than the regression-adjusted difference of -7.9, although both are large and statistically significant. Furthermore, the estimated

_____

[11]Appendix A discusses the sensitivity of our results to the possibility of incomplete billing by some agencies. In short, we find no evidence that incomplete billing is responsible for the large effects of prospective payment reported in this chapter.

TABLE III.9

UNADJUSTED ESTIMATES OF THE IMPACT OF PER-EPISODE PAYMENT ON THE NUMBER OF VISITS IN FIRST 120 DAYS
(Agencies Weighted Equally)

	Control Group Mean	Treatment Group Mean	Unadjusted Impact[a] (p-value)
Total Visits	45.03	33.67	−11.36 (.00)
Skilled Nursing Visits	21.71	16.34	−5.37 (.00)
Home Health Aide Visits	15.60	11.26	−4.34 (.02)
Physical Therapy Visits	5.63	4.69	−0.94 (.10)
Occupational Therapy Visits	1.01	0.62	−0.39 (.02)
Speech Therapy Visits	0.30	0.29	−0.01 (.87)
Medical Social Worker Visits	0.78	0.47	−0.31 (.00)

SOURCE: Medicare Claims Data.

Sample Size: 25,561 episodes in 44 treatment agencies; 25,130 episodes in 41 control agencies. Episodes are weighted so each agency has equal weight.

[a]The p-values are based on standard errors adjusted to account for the effects of clustering and weighting.

treatment-control difference in means for occupational therapy visits is statistically significant (p = .02), whereas the regression-adjusted difference is not (p = .12). Thus, despite the random assignment of the agencies to the treatment and control groups, the estimates are heavily dependent on the regression model, so the model specification is important.

The reason for the unexpected large difference between the regression-adjusted and unadjusted estimates is the sizable difference between the treatment and control groups in the predemonstration practice pattern variable. This variable measures the average number of visits per episode rendered by the agency relative to the average for all agencies during the predemonstration period, controlling for case-mix differences. As Table II.3 shows, the average value for this variable is materially smaller for treatment agencies than for control agencies. Because it is an extremely powerful predictor of visits per episode in the demonstration period, a significant portion of the observed treatment-control difference in mean visits per episode is attributable to the preexisting difference in practice patterns.[12] Thus, controlling for the fact that treatment agencies had lower-use practice patterns prior to the demonstration is important to avoid overestimating the reduction in visits attributable to prospective payment. The sizable treatment-control differences on a few other agency characteristics (size and location) had little effect on the estimated treatment-control difference in visits per episode, because they were less powerful predictors than the predemonstration practice pattern variable. Furthermore, we are relatively unconcerned about the possible omission of other agency characteristics from the regression model, since the effects of these characteristics on home health service use would operate through a predemonstration practice pattern, for which we control.

---

[12]The coefficient on this variable in the total visits regression (presented in Appendix Table B.1) was 24.5, with a t-statistic of 7.5. Since the difference in the treatment and control group means for this variable is about 0.16 (Table II.3), failure to control for this variable leads to a upward bias of roughly four visits when comparing unadjusted treatment-control means.

2. Weighting Agencies by Share of Episodes

Although giving agencies equal representation in the estimates is the most appropriate approach, impact estimates may be skewed by anomalous mean values of the dependent variable for one or two small agencies. The large weight given to episodes provided by these agencies makes the impact estimates particularly sensitive in turn to outlier values among the agencies' patients. An alternative approach is to weight observations so that each agency is represented in proportion to its share of the total number of episodes delivered by all demonstration agencies during their first demonstration year.[13] This approach gives far greater influence to agencies with large numbers of episodes than to agencies with few episodes. While the estimates obtained under "equal share" weighting are not sensitive to outlier values of small agencies and their patients, they are heavily dominated by the behavior of the largest agencies. We assess the sensitivity of our findings to the use of equal representation weights by comparing impact estimates for key outcomes under the two sets of weights.

The estimated effects of prospective payment obtained when weighting episodes by their agency share of total episodes are generally very similar to those obtained when giving each agency equal representation (Column 2 of Table III.10). While the magnitude of some impacts differs, there is no change in their statistical significance or direction. The largest difference observed is for the estimated impact on the number of home health aide visits: -1.7 visits per episode using share-of-episodes weights, but -2.9 visits using agency-equal weights. Overall, the estimated impact on the number of visits per episode when weighting by episode share is -7.1 visits, compared to -7.9 visits when weighting agencies equally.

---

[13]See Chapter II for a discussion of how this weight has been constructed.

TABLE III.10

ROBUSTNESS CHECKS FOR THE PRINCIPAL FINDINGS ON THE IMPACT OF
PER-EPISODE RATE SETTING DURING THE FIRST 120 DAYS

	OLS (Main)	OLS Weighted by Agency Share of Episodes	OLS Subsample Model (For Outliers)	Log-Linear Model (For Outliers)	Tobit Model (Censoring)
	-1	-2	-3	-4	-5
Total Visits	−7.85	−7.06	−7.90	−8.42	NA
(p-value)	(.00)	(.00)	(.00)	(.00)	
Skilled Nursing Visits	−3.97	−3.75	−3.87	NA	−3.39
(p-value)	(.00)	(.00)	(.00)		(.00)
Home Health Aide Visits	−2.86	−1.73	−3.25	NA	−1.75
(p-value)	(.01)	(.06)	(.00)		(.05)
Physical Therapy Visits	−0.55	−0.82	−0.62	NA	−0.39
(p-value)	(.29)	(.20)	(.14)		(.27)
Medical Social Worker Visits	−0.29	−0.32	−0.33	NA	−0.19
(p-value)	(.01)	(.00)	(.00)		(.04)
Episode Length	−9.78	−9.06	−10.26	NA	−11.01
(-p-value)	(.00)	(.00)	(.00)		(.00)
Sample Size					
Treatments (44 Agencies)	25,561	25,561	23,054	25,561	25,561
Controls (41 Agencies)	25,130	25,130	22,919	25,130	25,130

SOURCE: Medicare Claims Data.

NA = not applicable; OLS = ordinary least squares.

3. Effects of Outliers

To investigate whether our estimates have been unduly influenced by outliers (at the agency or episode level), we reestimated the regressions on visits per episode under two alternative specifications. First, we dropped from the regression models four treatment and four control agencies: those with the two highest and two lowest values in each group. Second, we respecified the dependent variable as the logarithm.[14]

The estimated impacts of prospective payment under these alternative models are very similar to those found in the main results (Columns 3 and 4 of Table III.10). For each of the outcome measures, the estimated impact changed little in size or statistical significance when the outlier agencies were dropped. The excluded agencies with large outlier values had total visits per episode of 56.8 and 71.6 for the treatment group and 77.0 and 78.7 for the control group. Agencies excluded because of the smallness of their mean number of total visits included treatment agencies with means of 12.7 and 13.0 visits per episodes and control agencies with 19.2 and 23.2 visits. These compare to overall means of 45 visits per episode for control group agencies and 34 visits per episode for the treatment agencies.

The estimated effect obtained by using the logarithm of total visits during the episode as the dependent variable (-8.4 visits) was also quite similar to the -7.9 visits estimate obtained from the regression. Both estimates were significantly different from zero.

---

[14]Because of the shortcomings of the logarithmic specification, especially for dependent variables with many zero values (see Chapter II), we use it only to examine the impacts on total visits.

4. Effects of Censoring

Several of our findings may be subject to bias due to censoring of the dependent variable.[15] All the outcomes associated with the number of visits (by type) reflect cases of left-hand censoring (at zero), while the episode length variable reflects a case of right-hand censoring (at 120). The statistical properties of these two forms of censoring are closely related, and, in principle, Tobit models control effectively for the possible bias introduced by both.[16]

The estimates from the Tobit models suggest that censoring has introduced no significant bias into our estimates (Column 5 of Table III.10). In our comparison of the OLS and Tobit results, the statistical significance and direction of all our impacts are consistent and the magnitudes are very similar. The only difference of any magnitude is for the estimated impact on home health aide visits: -3.1 visits using OLS regression and -2.1 using a Tobit regression.

5. Influence of Unobserved Variables

A final question is whether our results may be affected by unobserved differences between treatment and control agencies which we cannot control for through our demonstration period regression models. If, for example, treatment agencies have unobserved factors (local market conditions, differences in management) that lead them to provide relatively fewer visits per episode than control agencies in absence of the demonstration, our estimated demonstration impacts would be overstated.

---

[15]Under censoring, the dependent variable reflects a combination of discrete and continuous variables. Tobit models account for this censoring by estimating the parameters of a single model that predicts both the probability of using a service and the expected amount of the service.

[16]We report the prospective payment impacts using OLS specifications in the main part of the chapter because the statistical properties of the Tobit model can make its estimates unreliable. To the extent that the estimated impacts using Tobit and OLS are similar, however, it strongly suggests that our findings are not affected by the censoring or truncation of the dependent variable.

A straightforward approach to investigating the influence of unobservables on our findings is to examine the impact of an agency's treatment status on visits per-episode using data from the *predemonstration*. After controlling for observed characteristics available in the predemonstration period, we would expect an agency's treatment status to have no effect on visits per-episode in the predemonstration unless there were unobserved factors correlated with treatment status that also affect total visits.[17] The results from this regression (not shown) indicate that treatment status actually has a positive and significant effect of about 1.4 visits per episode in the predemonstration period. This suggests that our main findings may actually understate the influence of prospective payment on the use of services. Such an inference is not conclusive, however, since the lack of detailed case-mix data from the predemonstration may also explain some or all of this result.

An additional approach to investigating the effects of unobservables on our results is to examine how agencies changed their visits per episode between the predemonstration and the demonstration period. In particular, if the number of treatment agencies showing a decline in visits per episode is the same as than the number of control agencies, it suggests that our demonstration period analysis may in fact overstate the effects of the payment method on use.[18] However, this comparison (shown in Figure III.2) strongly confirms our main finding that prospective payment significantly reduced visits per episode. Most of the treatment group agencies show a decline in average visits per episode, relative to the base-year values, with the reduction being substantial for most

---

[17]We do not have detailed case mix data (ADLs, medical history, and so forth) for episodes during the predemonstration period. Thus, our control variables are limited to agency-level and area characteristics and basic demographic characteristics such as age and gender.

[18]The predemonstration data reflect all episodes that began in the first eight months of the fiscal year prior to when an agency entered the demonstration.

FIGURE III.2

DISTRIBUTION OF THE PERCENTAGE CHANGE
IN THE NUMBER OF VISITS PER EPISODE
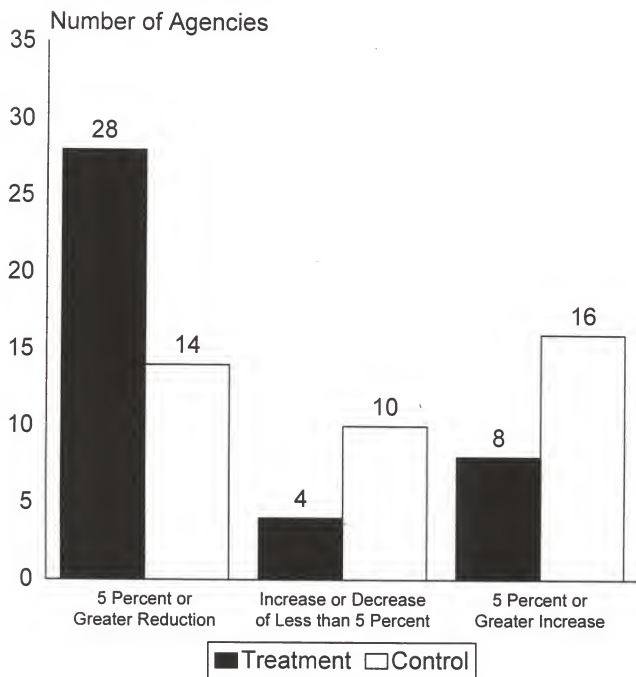
Source: Medicare Claims Data.

agencies (20 percent or more; not shown). Conversely, most control agencies exhibited either an increase or no change between the predemonstration and demonstration years.

6. Summary of Robustness Checks

In sum, the various robustness checks confirm the stability of our impact estimates. The estimated effects of prospective payment on visits per episode are large regardless of the how the episodes are weighted or what statistical model was used. The results are not driven by a few agencies that had exceptionally high or low visits during the period. The comparison of each agency's average visits per episode during the demonstration with its own predemonstration mean value confirms that nearly all treatment group agencies are lowering visits per episode in response to prospective payment incentives, whereas most control agencies exhibit little change in this measure over the same time interval.

IV. SUMMARY AND CONCLUSIONS

This preliminary analysis of demonstration impacts on the use of Medicare home health services provides HCFA with critical information for the design of the prospective payment system that is mandated for implementation in 1999. Our findings appear to be quite robust, suggesting that they provide a good guide for policymakers, despite their preliminary nature.

A. KEY FINDINGS

We find very strong evidence that prospective payment reduced the number of home health visits by a substantial margin. Prospectively paid agencies rendered approximately 17 percent fewer visits per episode than the control agencies. This reduction involved large and statistically significant declines in the number of visits by skilled nurses, aides, and medical social workers, as well as large but statistically insignificant declines in the number of visits by physical therapists and occupational therapists. The effect on speech therapy visits, which accounted for less than one percent of all visits, is positive and not statistically significant.

Agencies appeared to reduce all major types of visits per episode, in roughly the same proportions. Demonstration effects on the proportion of total episode visits for skilled nursing, aide care, therapies (combined), and medical social services were all very small and statistically insignificant. We find that prospective payment has no major effect on the probability of receiving *any* visits of a particular type, except for occupational therapy, for which the probability drops by about one-third. Skilled nursing and aide visits each were reduced by about 17 percent for recipients of these services, the same reduction observed in total visits. Estimated effects on the less used services (therapies and medical social) for recipients of these services varied somewhat. Overall,

however, it appears that agencies found ways to reduce all types of home health services, except speech therapy, by substantial amounts.

We also find that prospective payment reduced the average length of home health episodes. Consistent with the reduction in the number of visits provided, we estimate that prospective payment reduced the length of episodes within the "at-risk" period by about 10 days, a drop of nearly 15 percent relative to the mean for control group agencies. Prospective payment reduced the proportion of episodes lasting over 120 days by about 30 percent, lowering it from about 35 percent to 25 percent.

The demonstration effects on the number of visits per episode are fairly similar across subgroups of agencies defined by their for-profit status, facility-based/freestanding status, and size, but differ significantly with their predemonstration practice patterns. Treatment group agencies that delivered a higher-than-average number of visits per episode in the predemonstration period reduced visits by over twice as much as those with low-use prior practice patterns. The reductions for both subgroups were significantly different from zero, however, and the reduction as a percentage of mean visits was substantial for both. Impacts on the mix of visits per episode and episode length do not appear to vary with the agency characteristics examined.

B. POLICY IMPLICATIONS

Our findings suggest that prospective payment for home health agencies could lead to potentially sizable savings in costs, depending upon how the payment rates are set. The 17 percent reduction in visits does not necessarily imply similar size reductions in costs, however, because agencies' cost per visit may increase if their total volume of visits declines. Nonetheless, if savings of roughly this magnitude are achievable once agencies adjust to a smaller scale or consolidate, the potential exists both for HCFA to save and for agencies to prosper. This could be achieved by

setting payment rates at a proportion (for example, 90 percent) of the average cost per episode under cost-based reimbursement, much as is done in the Medicare risk program for payments to HMOs.

The finding that nearly all types of visits appear to be reduced suggests that it could be beneficial for HCFA to design its future prospective payment system in a way that gives agencies the flexibility to chose which services to decrease. Recently, in the Balanced Budget Act of 1997, Congress restricted coverage for skilled nursing for patients needing blood drawn. This effort to reduce home health utilization focused on decreasing the use of skilled nursing and home health aide services, which is logical because these services are used most frequently and hence are thought to provide the greatest opportunities for cost savings. However, this analysis shows that given the right incentives, agencies may make substantial reductions in therapy and medical social worker visits, which account for about 17 percent of all visits in the first 120 days of an episode. Therefore, a prospective payment system that encompasses all services rendered could be beneficial to the government.

The observed reduction in episode length is also potentially important for designing the future system. In recent testimony before the Subcommittee on Oversight and Investigations, the Director of Health Financing and Systems Issues for the U.S. General Accounting Office, Dr. William Scanlon, argued that in the design of a payment system for home health care, "the episode should be long enough to capture the care typically furnished to patients," and noted that an at-risk period of 150 days would be required to cover 82 percent of all episodes (Scanlon 1997). Based on our estimates of the decrease in long episodes, it appears that HCFA could design a system based on a 120-day at-risk period and still encompass 80 percent of all Medicare home health patients. Thus, HCFA may wish to consider using a shorter at-risk period than the experience under cost-based reimbursement would suggest.

C. LIMITATIONS OF THE ANALYSIS

Although we are confident that the estimates reflect real effects of prospective payment on home health agency behavior, four potential limitations should be kept in mind. First, there is some chance that the estimated treatment-control difference in visits per episode may be overstated due to under-reporting by treatment group agencies. Agencies paid prospectively have little incentive to be sure that claims for all services provided during a given episode have been submitted, once the initial bill that generates the per episode payment has been sent in. We discovered two treatment group agencies for which there were major billing irregularities and excluded them from the analyses presented. However, if several other treatment group agencies fail to submit a moderate fraction (say, 10 percent) of their second and subsequent bills for patients, our estimates would overstate the impact of prospective payment on service provision. Our preliminary investigations suggest that billing irregularities are relatively rare for most agencies, but we are continuing to explore this issue.

The second concern is whether the results can be generalized to other agencies. As in any study in which the participants are volunteers, demonstration agencies may be those best able to respond to the incentives of the demonstration. If this were true, the impacts of prospective payment if instituted nationally are likely to be smaller than those estimated here. On the other hand, our subgroup analysis suggests that if agencies participating in the demonstration previously had lower-than-average-use practice patterns, reductions in visits under a national program could be larger than those estimated here. We have no evidence that participants are unrepresentative of agencies nationally, but will explore this issue further in the final report.

We are also concerned that the results may not provide a reliable guide to how agencies would behave under a national program, because of the ways in which such a program would likely differ from the demonstration program. Under a national program of prospective rate setting, agencies

would not be protected from incurring financial losses, which could compel some to respond more aggressively to prospective payment. Furthermore, the per-episode rate paid to an agency would probably not be based on its own prior cost per episode, but on a regional or national average, greatly increasing the potential for losses for agencies with high-use practice styles and those with high costs per visit. The financial pressure on agencies would be intensified if the rate were based on the episode cost of an "efficient" agency or if HCFA sets the per-episode payment at some percentage (for example, 90 percent) of the average national per-episode cost, in order to share in the savings that per-episode payment will generate. Indeed, the Balanced Budget Act of 1997 requires that the future prospective payment system be budget neutral to the current Interim Payment System (which introduced substantial cuts) *minus* 15 percent. The additional financial pressures will likely produce prospective payment effects even greater than those observed here.

Finally, we must point out that these results are preliminary, which limits the inferences that can be drawn on several dimensions:

- The analysis covers only the first 8 to 15 months of the three-year demonstration period for participating agencies. Impacts may change as agencies become more adept at and comfortable with finding ways to reduce the number of visits provided without affecting the quality of care.

- These results provide no information on the consequences of the reduced level of services for home health recipients or their families (for example, quality of care, access, and caregiver burden may be affected).

- The estimated reductions in visits do not necessarily translate into proportionate reductions in the cost per episode, because the cost per visit may increase if an agency's total volume declines. Program effects on home health costs for the period after the first 120 days of the episode must also be examined. Costs for other Medicare-covered services may be affected as well.

- We do not yet know whether and how the estimated reductions in home health services tend to vary with patient characteristics.

These issues will all be addressed in future reports, as the required data become available.

D. CONCLUSION

Despite the limitations of these preliminary estimates, the conclusion is clear: prospective payment seems to be a promising alternative to the present payment system. If our future analyses confirm our belief that the impacts are not overstated because of underreporting, and no adverse effects on the quality of care or access to care are observed, prospective payment may be a viable mechanism for Medicare to reverse the pattern of large increases in home health care expenditures.

REFERENCES

Brown, Randall, Barbara Phillips, Christine Bishop, Amy Klein, Grant Ritter, Craig Thornton, Peter Schochet, and Kathleen Skwara. "The Effects of Predetermined Payment Rates for Home Health Care." Princeton, NJ: Mathematica Policy Research, Inc., December 1995.

Duan, N., Willard G. Manning, Jr., Carl N. Morris, and Joseph Newhouse. "A Comparison of Alternative Models for the Demand of Medical Care." *Journal of Business and Economic Statistics*, vol. 1, no. 2, April 1983.

Health Care Financing Review. *Medicare and Medicaid Statistical Supplement*, 1996.

Phillips, Barbara, and Rachel Thompson. "Transition Within a Turbulent System: An Analysis of the Initial Implementation of the Per-Episode Home Health Prospective Payment Demonstration." Draft report. Princeton, NJ: Mathematica Policy Research, Inc., February 1997.

Phillips, Barbara R., Randall Brown, Valerie Cheh, Amy Klein, Jennifer Schore, Robert St. Peter, and Craig Thornton. "Evaluation of the Per-Episode Home Health Prospective Payment Demonstration Design." Report submitted to the Health Care Financing Administration. Princeton, NJ: Mathematica Policy Research, Inc., September 1995.

Prospective Payment Assessment Commission. *Report and Recommendations to Congress*. Washington, DC: PPAC, March 1, 1996.

Scanlon, William. "Success of Balanced Budget Act Cost Controls Depends on Effective and Timely Implementation." Testimony Before the Subcommittee on Oversight and Investigations, Committee on Commerce, House of Representatives, October 29, 1997.

Schore, Jennifer. "Regional Variation in the Use of Medicare Home Health Services." In *Persons with Disabilities: Issues in Health Care Financing and Service Delivery*, edited by Wiener et al. Washington, DC: The Brookings Institution, 1995.

Teplitsky, Sanford, V., and Mary Ann Janson (eds.). *Home Health & Hospice Manual: Regulations and Guidelines*. Owings Mills, MD: National Health Publishing, 1985-1992.

APPENDIX A

DATA QUALITY

As noted in Chapter II, we discovered during our analysis that two agencies were not submitting a large portion of their bills for services rendered during the episode. We identified these agencies (both treatments) because they showed extremely large reductions in visits per episode from the predemonstration to the demonstration, and they also had among the highest rates of one or two visit episodes. Because this nonbilling leads the observed number of visits per episode to be smaller than the actual number provided, any analysis that included these two (treatment) agencies would overestimate the reduction in visits per episode brought about by the demonstration payment method. For this reason, we excluded both these agencies from our main analysis.

The discovery of nonbilling by these two agencies raised immediate concerns that other agencies may not be submitting bills as well but that their behavior was simply less noticeable. To the extent that billing problems are more widespread, its effect on our impact estimates will depend on whether the problem is random across treatment and controls. In cases where missing or erroneous bills are caused by factors unrelated to the payment method (delays in getting physician signatures, untimely bill submission, and so forth), they are likely to exist about equally across both treatment and controls and therefore have little influence on our results. However, one of the two agencies that we excluded from our analysis clearly indicated that its behavior was a direct result of the demonstration. Namely, because it was paid per episode and upon acceptance of the initial bill, the agency did not believe it was in its best interest to incur billing costs for visits rendered later in the 120-day at-risk period.[1] To the extent that other treatment agencies pursued a similar strategy, our impact estimates may be overstated.

---

[1]Because of the complicated nature of the cost-reporting process and the calculation of losses, there is some financial advantage to bill completely. However, not all agencies understood the advantage at this point in the demonstration.

This appendix describes the billing issues in detail and provides a sensitivity analysis of our results to this problem.

A. IDENTIFYING AGENCIES WITH MISSING OR ERRONEOUS CLAIMS

To measure the extent and source of any billing problems among the demonstration agencies, we initially conducted two data analyses. The first examined bill records and checked for the frequency of missing interim bills. Using episodes that were at least 60 days long, we identified those episodes that included only an initial bill and a discharge bill, and we compared the proportion of these episodes across agencies. The second analysis used Medicare claims data and the quality assurance (QA) data. In this analysis, we compared the bill through date of the last Medicare bill of the episode with the last day of service as recorded on the QA forms. If agencies failed to submit all their bills, then the last day of service from the QA form would be after the discharge date obtained from the bills.

Each of these analyses had significant limitations that hampered our ability to identify agencies with possible billing problems. In the case of searching for missing interim bills, we found that submission of only the initial and final bills was commonplace in many agencies, often involving more than 20 percent of their episodes regardless of treatment status. This made it difficult to distinguish agencies that had truly failed to submit bills from those who had not. In the case of analyzing the QA data, we had information only on episodes starting after April 1996--several months after the last agencies had enrolled in the demonstration. As a result, these data included only a small fraction of many agencies' episodes, leading us to question whether they were representative of all episodes. In addition, early in the demonstration, not all agencies submitted QA data systematically, leading to an even smaller (and potentially more nonrepresentative) sample of episodes within each agency. Finally, we discovered that for agencies using telephone contacts in

lieu of visits, the last date of service (the telephone contact) may be after the discharge date (the last visit date). Because treatment agencies have the financial incentive to make telephone contacts, they would likely have higher frequencies of the last date of service following a discharge date.

Despite these limitations, we were able to identify nine agencies whose pattern of intermittent billing and mismatching discharge dates were suspiciously out of the ordinary. (These nine agencies included the two agencies that we had already excluded from the main analysis.) In order to collect more detailed information, we contacted the billing staff of these agencies and requested information about their practices and any problems with submission of bills. The results from the interviews were as follows: of the nine agencies, eight indicated that they did have trouble billing. (The one agency that did not report trouble billing was very small, giving us little data on which to base our "suspicions.") Of the remaining eight agencies indicating that they had submitted incomplete bills, four agencies (including one of the two excluded form our main analysis) stated that they did not bill completely because of the new payment system. The agencies cited two general reasons for not billing. In one case, the agency's software did not allow for visits to be billed at $0 cost. In the other three cases, the agencies chose to completely bill only for services that would generate an additional payment. In all four cases, some (but not all) of these agencies' interim bills were submitted.

Because it was possible that other agencies had failed to bill completely but we had simply not been able to identify them, we decided to contact all the agencies in the demonstration and request similar information about their billing practices. From these contacts, we identified one additional treatment agency that had a problem with billing directly related to the payment method. In this case, the agency submitted claims for all visits, but if a claim was "kicked back" for any reason, only the initial episode claim or claims with outlier visits were resubmitted. In addition, a total of 21

other agencies--12 treatments and 9 controls--reported some amount of incomplete bills, but for reasons unrelated to the demonstration. The most common reason was that the local doctor had refused to sign the plan of treatment, making it impossible for the agency to bill Medicare for the service.

B. THE SENSITIVITY OF OUR RESULTS TO BILLING PROBLEMS

To understand how our estimates may be affected by incomplete bills among agencies, we reestimated the impact of prospective payment on total visits using four additional samples. The samples created are:

1. All (87) agencies for which we have demonstration data, including the two agencies excluded from the main analysis. Results from this regression provide an upward bound on our impact estimate, since we know these two agencies have particularly low levels of (observed) visits per episode.

2. All agencies, excluding the five (treatment) agencies reporting that they failed to submit bills for reasons related to the new payment system. Results from this regression may understate the actual impact estimate, since it excludes agencies that may have substantially responded to the demonstration incentives (that is, cut visits) in addition to failing to submit bills. Nevertheless, the impact estimate provides an indication of how incomplete billing by treatment agencies may affect our main findings.

3. The same set of agencies as in (2) but also excluding the second agency with major billing errors that we omitted from the main analysis. (This agency had billing problems unrelated to the demonstration payment.) Results from this regression provide an indication of whether our estimate in (2) is greatly affected by the inclusion of an outlier agency.

4. The same set of agencies as in (2) but further excluding the 21 agencies (12 treatment and 9 control agencies) reporting problems submitting bills unrelated to the demonstration. This regression should provide the most conservative estimate of the demonstration impact, since it excludes any agency that reported a problem with billing. In this sense, the regression result may be viewed as a lower bound on our main impact estimate.

The regression results (shown in Appendix Table A.1) show no qualitative difference in the estimated impacts of the demonstration using any of the different samples. Not surprisingly, the largest impact estimate (-8.7 visits per episode) is obtained using the full sample, which includes the two treatment agencies excluded from our main analysis for significant incomplete bills. This estimate, however, is less than one visit larger than the impact estimate from our main analysis, -7.9 visits per episode. The smallest estimate (-5.8 visits per episode) is obtained using the sample that excludes all agencies with incomplete bills related to the demonstration payment method, as well as the additional outlier agency that we excluded from the main analysis. This impact is about two visits smaller than our impact reported in the main results, but it is still very large and significant. As a percentage of the control group mean (45 visits per episode), it reflects a reduction in visits per episode of nearly 13 percent. Moreover, based on the standard error of this estimate (1.6 visits), it is not possible to reject the hypothesis that it is the same effect found in the main analysis. Finally, the estimated impact when excluding all agencies reporting any billing problems is -6.8 visits per episode, a difference of only one visit from the impact estimate reported in the main results.

From this analysis, we conclude that the existence of incomplete bills by agencies is problematic; however, it is not responsible for the large impact of the demonstration payment method on the provision of services. In future reports, we will continue to monitor the existence and reasons for nonbilling by agencies and provide a careful investigation of the sensitivity of our main findings to this problem.

TABLE A.1

THE IMPACT OF PER-EPISODE PAYMENT ON VISITS PER EPISODE
UNDER VARIOUS SAMPLES DEFINED BY BILLING PROBLEMS
(Agencies Weighted Equally)

	Impact[a] (p-value)	Number of Agencies
Main Analysis Sample	-7.85*** (.00)	85 (44 treatment)
Sample 1: All Agencies	-8.67*** (.00)	87 (46 Treatment)
Sample 2: All Agencies, Excluding Five with Incomplete Bills Due to the Payment Method	-6.15** (.00)	82 (41 (Treatment)
Sample 3: Sample (2), Excluding an Additional Agency Excluded from the Main Analysis Sample	-5.91*** (.00)	81 (40 Treatment)
Sample 4: All Agencies, Excluding Any Agency Reporting Some Level of Incomplete Bills for Any Reason	-6.83*** (.00)	60 (28 Treatment)

SOURCE: Medicare Claims Data.

[a]The estimates, obtained from ordinary least squares regression models, represent treatment-control differences in the mean number of visits per episode. The p-values are based on standard errors that account for the effects of clustering and weighting.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.
***Significantly different from zero at the .01 level, two-tailed test.

APPENDIX B

**COEFFICIENT ESTIMATES FROM THE REGRESSION
ON TOTAL VISITS PER EPISODE**

The complete list of coefficients from the regression on visits per episode (shown in Appendix Table B.1) indicates that a number of patient- and agency-level characteristics have significant effects.

Nearly all of the patient characteristics associated with case mix (including ADLs, medical conditions, and length of prior hospitalization) have significant effects in the expected direction. Among these case-mix variables, those having the largest impact are wound care (11.2 visits) and bathing (8.1 visits), as well as diabetes (7.6 visits) and cardiovascular accident (7.3 visits). Most of the demographic variables (unrelated to case mix) also have a significant effect on visits. All else equal, older age groups (75 and over) have 3 to 3.5 more visits per episode than those 65 to 74, while those with age as original reason for entitlement have 3.5 fewer visits than those whose original reason for entitlement is disability. Finally, whites have about 4.5 fewer visits than nonwhites.

Aside from treatment status, only a few of the agency-level variables have significant impacts on visits. By far the largest of these is the agency practice pattern.[1] Given a one standard deviation increase in the agency practice pattern (about 0.31), for example, the number of visits per episode increases by about 7.6 visits. The state in which a given agency is located also has a significant effect on visits. Controlling for other agency characteristics, agencies in Massachusetts and Florida provide far more visits per episode (11.4 and 8.8, respectively) than those in California (the base category); however, there is no significant difference among agencies in California, Texas, and Illinois. Somewhat surprisingly, there is no significant difference in the provision of visits depending on whether an agency is nonprofit or proprietary, relatively small or large, or freestanding or hospital-based.

---

[1]This variable reflects the case-mix-adjusted ratio of the average visits per episode in a given agency and the average visits per episode in all agencies. For further detail on this variable, see the data section of Chapter II.

COEFFICIENT ESTIMATES FROM OLS REGRESSION ON TOTAL VISITS PER EPISODE
(AGENCIES WEIGHTED EQUALLY)

	Number of Visits per Episode	(p-value)
Intercept	-7.92	(.040)
Agency Received Prospective Payment	-7.85	(0.00)
Age		
Younger than 65	-0.53	(0.82)
75 to 84	3.56	(0.00)
85 or older	3.05	(0.04)
Female	0.18	(0.87)
White	-4.47	(0.01)
Original Reason for Medicare: Old Age	-3.47	(0.07)
Medical Conditions		
Cancer	-2.14	(0.10)
Diabetes	7.65	(0.00)
Cerebrovascular accident (stroke)	7.28	(0.00)
Decubiti stage 3 or 4	6.08	(0.0)
Need for Complicated Wound Care[a]	11.21	(0.00)
Functional Limitations[b]		
Bathing	8.10	(0.00)
Eating	2.98	(0.01)
Dressing	3.35	(0.00)
Toileting	6.01	(0.00)
Transferring	3.38	(0.00)
Preadmission Location: Hospital	-4.67	(0.00)
Had Medicare for Less than Six Months	0.55	(0.84)
Practice Patterns	24.47	(0.00)
Hospital-Based Agency	0.21	(0.89)
For-Profit Agency	2.20	(0.24)
Chain Member	0.61	(0.78)
Agency Provided Fewer than 30,000 Visits in Base Year	3.58	(0.12)
Agency Located in Urban Area	-1.53	(0.67)

	Number of Visits per Episode	(p-value)
Agency State		
Florida	8.79	(0.00)
Illinois	2.20	(0.44)
Massachusetts	11.40	(0.00)
Texas	3.92	(0.11)
County-Level Means		
Number of nursing home beds per 100 persons over age 65	1.14	(0.10)
Number of physicians per 10,000 persons	-0.08	(0.56)
Hospital occupancy rate	0.18	(0.99)
Length of Hospital Stay During Two Weeks Before Home Health (Days)[c]	0.28	(0.28)
Any Skilled Nursing Facility Stay During Two Weeks Before Home Health	5.01	(0.00)
Total Medicare Part A Reimbursement During Six Months Before Home Health (In Thousands of Dollars)[e]	.16	(0.00)
Number of Episodes	50,691	

SOURCE: Medicare Standard Analytic Files 1993 to 1995.

[a] Patient has wound that requires soaking, irrigation, or debridement.

[b] Patient requires some human assistance with or does not participate in activity.

[c] A case-mix-adjusted index of the average number of visits received by an agency's patients relative to the average number provided by other demonstration agencies.

[d] If patient was not hospitalized within the two weeks before home health, days are set to zero.

[e] Includes reimbursement for inpatient, skilled nursing facility, hospice, and nondemonstration home health paid under Medicare Parts A and B.

  *Significantly different from zero at the .10 level, two-tailed test.
 **Significantly different from zero at the .05 level, two-tailed test.
***Significantly different from zero at the .01 level, two-tailed test.

APPENDIX C:

**CASEMIX ADJUSTMENTS DURING
THE DEMONSTRATION**

During the demonstration, agency payment rates will be adjusted for severity using the Home Health Utilization Groups (HHUGs) Classification system. This system, which was specifically developed for the demonstration, classifies patients into 1 of 18 mutually exclusive cells on the basis of the following information:

- Whether the patient has an intervening hospital stay in an acute-care hospital during the 120 days following home health admission

- Whether the patient was impaired in at least four of five activities of daily living (bathing, eating/tube feeding, dressing, toileting/elimination, and transferring) at home health admission

- Whether wound care was planned for the patient at home health admission

- Whether the patient was discharged from a hospital within the 14 days preceding home health admission

- Whether the patient had stage 3 or stage 4 decubitus ulcer at home health admission

- Whether the patient had cancer at home health admission that affected current treatment or personal-care needs

- Whether the patient has had a cerebrovascular accident that affected current treatment or personal-care needs

- Whether the patient had diabetes that affected current treatment or personal-care needs

As with payment rates, demonstration case-mix adjustment is based on the agency's base-period experience. Information on the characteristics listed above was collected on each patient admitted to each agency in the 90 days ("base quarter") before it began to operate under the demonstration. The same information was collected throughout the demonstration in the "remarks" section of the Uniform Bill, HCFA form 92 (UB-92).

Using the base quarter information, Abt calculated a base quarter index for each agency as follows. First, Abt calculated a category weight for each of the 18 casemix groups by dividing the

average cost in each group by the agency's overall average cost.[1] Abt then multiplied each category weight by the percentage of episodes that fell into each casemix category during the base quarter, and summed across all 18 categories. The result is a base quarter index for each agency, which is always equal to 1.

To obtain the annual case-mix adjusted rates, Abt calculated a similar index for the demonstration years by multiplying the category weight from the base year times the percent of episodes that fell into each case mix category during the relevant year, and summed across all 18 categories.

The demonstration year index for each agency is then divided by the agency's base quarter index; and is multiplied by the agency's base year episode rate. This results in the casemix adjusted episode rate

---

[1] If an agency did not have any observations in a particular category during the base quarter, the average cost for that casemix group calculated across all demonstration agencies was used instead of an agency-specific cost.